



UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA IN
STATISTICA APPLICATA

CICLO XXVII

COORDINATORE Prof. Corradi Fabio

Identifiability of Discrete
Hierarchical Models with One Latent
Variable

Settore Scientifico Disciplinare SECS-S/01

Dottorando

Dott. Biondi Luigi

Tutore

Prof. Marchetti Giovanni Maria

Coordinatore

Prof. Corradi Fabio

Anni 2012/2015

Contents

1	Introduction and Outline	1
1.1	The problem of identifiability	1
1.2	Outline of this work	2
1.3	Historical Overview	3
2	Latent Variable Models	6
2.1	Introduction to Latent Variables	6
2.2	Graphical Models	8
2.2.1	Markov Properties	11
2.3	Log-linear Models	12
2.3.1	Hierarchical Log-linear models	16
2.3.2	Hierarchical log-linear models with one hidden binary variable	17
3	Identifiability	20
3.1	Generic and Local Identifiability	20
3.1.1	Finite and local identifiability	23
3.1.2	The Jacobian Matrix	25
3.1.3	Generic identifiability	26
3.2	Geometric Interpretation	27
3.3	Identifiability for Poisson and the multinomial models	29

4	Identifiability for Hierarchical Log-linear Models with one hidden variable	32
4.1	Identifiability of graphical models with one binary latent Variable	32
4.2	Non graphical Hierarchical Log-Linear Models	35
4.2.1	Identification for hierarchical models with interactions up to the second order	38
4.3	Extension to the case of observed variables with more than two levels	49
5	Identifiability for models with higher order of interactions	52
5.1	Generic Identifiability for models with III order interactions . . .	52
5.1.1	Models with 4 observed variables	53
5.1.2	Models with 5 observed variables	55
5.2	Theorem	58
5.2.1	Consequences of the characterization of generic identifiability for models with interactions up to III order	59
5.3	Generic Identifiability for models with interactions of order 4 or more	60
5.4	Behaviour of the model near the critical points	65

List of Figures

2.2.1 An example of undirected graph: the <i>star graph</i>	9
2.2.2 An example of <i>tree</i>	10
2.3.1 Graph G of the example 2.3.1	16
2.3.2 A possible graph for a model for the Example 2.1.2	19
3.1.1 Local Identifiability	24
3.2.1 \mathcal{M}_G is not locally identifiable	28
4.1.1 A model that is graphical with respect to graph G is generically identifiable even if condition (i) does not hold.	34
4.2.1 If at least one of these two graphs is a subgraph of \overline{G}^O , then \mathcal{M}_G^2 is locally identifiable	39
4.2.2 G^O and \overline{G}^O	43
4.2.3 Two complete components	45
5.1.1 Proposition 4.2.6 cannot be extended to higher orders.	52
5.1.2 Direct and complementary graphs for the only model with 4 ob- served variables without a 3-clique in the complementary graph that is not overparametrized when we include III order interac- tions.	53
5.1.3 Direct and complementary graph of the observed variables for Model A	55
5.1.4 Overparametrized graphical submodel of model A	56

5.1.5 Non-identifiable graphical submodel of model A	56
5.1.6 Direct and complementary graph of the observed variables for model B	57
5.4.1 Graph of the observed variables for the example.	65
5.4.2 Behaviour of the model in a neighbourhood of a critical point . .	66

Chapter 1

Introduction and Outline

1.1 The problem of identifiability

Identifiability is a crucial property in statistics, as it is a primary requirement for models in order to be correctly estimated.

A statistical model is a collection of probability distributions. A model can be defined by a function from a set of parameters: in this case it is called *parametric model*, and the function itself is called parametrization map. Identifiability is a “good property” for the parametrization map: a model is identifiable if this function is one to one, that is, each probability distribution in the model is given only by one parameter. If identifiability fails, it means that there is not a unique way to locate a specific probability in the model. When there are multiple choice of parameters for every single distribution in a model, interpretation can be very difficult.

The simplest cause of non-identifiability is *overparametrization*, that is, when it is possible to describe the same model with a strictly smaller set of parameters. This problem can be solved by imposing some constraints on the parameters, for example by setting to zero some of them. However, there are other, more subtle, possible cases: in fact other weak form of identifiability have been introduced in literature such as local and generic identifiability. The objective of this work is to detect such non-trivial cases and explore the consequences of the

lack of identifiability for models with latent variables.

1.2 Outline of this work

We begin this work by giving a brief historical overview on the problem of the identifiability in the statistical literature.

In Chapter 2 we introduce the latent variable models, and we will give a few motivating examples in order to show their usefulness. Latent variable models make use of one or more non-observable random variables to explain the variability in the data. The archetypical model is the *latent class model*, where the observable variables are assumed independent when conditioned to the latent variable. It is therefore necessary to adopt a parametrization that allows to easily define conditional independencies. This role is taken by graphical log-linear models, of which we will give an essential overview in the rest of the chapter. As the name suggests, graphical models are models defined by a graph, using the edges between nodes to represent dependencies.

The log-linear parametrization is a handy way to parametrize those models.

In Chapter 3 we explore in detail different notions of identifiability:

- *global* identifiability, strongest condition, unattainable in most practical cases;
- *finite* and *local* identifiability, more relaxed conditions that retain most of the good properties;
- *generic* identifiability, a condition that allows the presence of *critical points* that possibly can give some problems in the estimation process.

These definitions are all connected to the jacobian matrix of the parametrization map: the goal of this entire work will be to establish conditions that guarantee

that this matrix has maximal column rank. A few examples are then given, in order to show the consequences of these conditions.

Chapter 4 presents the first original results of the thesis: first, we present the characterization given by Stanghellini and Vantaggi (2013, [26]) for the identifiability of graphical models. Then, we move further with a characterization of the identifiability for a different class of models: hierarchical models with interactions of order at most 2. This result is complete: we have found a simple necessary and sufficient condition for models with full rank matrices, based on the topology of the graphs encoding all the independences. It turned out that 5 observed variables are sufficient for achieving local identifiability in this class of models.

Finally, in Chapter 5, we present some results on hierarchical models with interactions of higher order. We give a general recursive procedure to assess if a log-linear hierarchical model is generically identifiable. Some simulation results suggest that generic identifiability holds when the number of observed variables is twice the maximal order of the interactions included in the model.

Most of the original research in this thesis has been done together with Barbara Vantaggi, whom I heartily thank.

1.3 Historical Overview

In 1997 a paper from Catchpole and Morgan introduces the concept of local identifiability as a weaker condition than the injectivity of the parametrization map. They prove a sufficient condition for local identifiability for Poisson and multinomial models for contingency tables, by studying the rank of the jacobian matrix of the parametrization map.

Stanghellini ([25], 1997) and Vicard ([27], 2000) found results on necessary and sufficient condition for the identifiability of single-factor models, by study-

ing the zero-elements of the concentration matrices. Those results are presented as conditions on the graphical structures defining those models.

In 2009 a paper from Allman, Matias and Rhodes ([3]) proved some results on the generic identifiability for models with one discrete latent variable. They introduce the Kruskal decomposition theorem as a tool for studying identifiability in presence of latent variables. Given a latent class model with a hidden variable with r hidden classes and any number of observed variables, they find that a sufficient condition for generic identifiability is that

$$\min(r, k_1) + \min(r, k_2) + \min(r, k_3) \geq 2r + 2,$$

where k_1, k_2 and k_3 are the cardinalities of an appropriate tripartition of the set of observed variables. As a corollary there is a sufficient condition for the identifiability of r -mixtures of products of p independent Bernoulli variables:

$$p \geq 2\lceil \log_2 r \rceil + 1.$$

In the same year, Drton showed some consequences of the presence of singular points for standard inferential procedures such that LRT tests. Since the limiting distribution of the likelihood ratio statistic depends on the geometry of the tangent cone in the true parameter, when this one lies in the set of singular points (and thus the tangent cone is possibly not convex) nonstandard distributions arise, like minima of possibly dependent χ^2 distributions.

In 2013 Stanghellini and Vantaggi found a complete characterization of the local identifiability for graphical models with one latent binary variable, based on the topology of the graph: the key conditions are the presence of a complete subgraph of order at least 3 in the complementary graph (a condition that is sufficient for the generic identifiability) and the existence of a *identifying sequence*

of subgraphs for each clique of the graph. Furthermore, for those graphs having as submodel a latent class model with 3 observed variables and such that, for some clique, there is no generalized identifying sequence, the space where the rank of jacobian is not full is determined. This subspace has null Lebesgue measure.

In 2014 Allman, Rhodes and Stanghellini and Valtorta ([4]) analyze generic identifiability for small bayesian networks with at most 4 discrete variables by using Kruskal decomposition, giving explicit identifiability procedures for DAGs. Moreover, they construct a parametrization map for a model with 4 observed variables and one hidden variable that is generically 4-to-1.

Chapter 2

Latent Variable Models

In this work we will examine in detail models where one or more variables are *latent*, that is, where some of the variables are non-observable. This chapter introduces in detail this framework and the reasons behind this choice of models.

2.1 Introduction to Latent Variables

As pointed out by Bartholomew et al. ([5]), there are many reasons to use latent variable models. First, they allow us to describe the information contained in the data with a much smaller set of variables, a desirable property as it improves our ability to see the underlying structure of the data. The presence of too many variables is sometimes cumbersome and it may hide some of the patterns in their inter-relationships, as we are unable to visualize them in more than our usual 3 spatial dimensions. Latent variables provide a good way to condense the informations hidden in data with many variables without too much loss.

However, latent variables are not just a technical tool: sometimes we are also needed to investigate the relationships between quantities that are in fact non-observable. For example, this happens in social sciences, when we may be interest in evaluating non-misurable quantities like quality of life or general intelligence (as in [6]). These concepts are hard to quantify in a rigorous, mathematical sense, and latent variable models provide a way to effectively describe them.

Example 2.1.1 (Agresti, [2]). Suppose we have data from k different pathologists who classified n slides on the presence or absence of some kind of carcinoma. For modeling interobserver agreement, we can make use a latent variable with $q = 2$ classes, one for subjects whose true rating is positive and one for subjects whose true rating is negative.

In this way, we express the joint distribution of the 2^k entries of the contingency table with the ratings as a mixture of two distributions, one for each true rating class.

Example 2.1.2. The National Job Corps Study was a randomized experiment performed back in the mid 1990s to evaluate the effect of participation in a large job training program for economically disadvantaged youths (see [7] for more details).

A random sample of eligible applicants was randomly assigned to the training program, while another random sample was assigned to the control group by denying the access to the program for 3 years. These groups were tracked and data of their employment status was gathered soon and at 12, 30, and 48 months after randomization.

This situation is analyzed in [19] by Mealli, Pacini and Stanghellini. The variables in the dataset are the following:

- binary smoking habits at 12th month (CIG12);
- binary employment indicators at 12th (W12), 30th (W30) and 48th month (W48).
- binary variables for the assignment and actual participation to the training program.

Only observations where all the outcomes and the treatment indicator are not missing (N=8291) are used.

Under the stable unit treatment value assumption (SUTVA: no interference between units nor different versions of the treatment, see [23]), let Z_i be a binary treatment assignment for unit i (0 for the control group, 1 for the treatment group). We denote by $D_i(z)$ the potential intermediate binary variable for unit i when assigned treatment z : in this case, the actual participation to the training program. The units under study can be stratified into the following four subpopulations, according to the value of the two potential indicators $D_i(0)$ and $D_i(1)$:

- 11 = $\{i : D_i(1) = D_i(0) = 1\}$ (*always-takers*),
- 10 = $\{i : D_i(1) = 1, D_i(0) = 0\}$ (*compliers*),
- 01 = $\{i : D_i(1) = 0, D_i(0) = 1\}$ (*defiers*),
- 00 = $\{i : D_i(1) = D_i(0) = 0\}$ (*never-takers*).

Because only one of the two potential indicators is observed, these four subpopulations are latent, in the sense that in general it is not possible to identify the specific subpopulation a unit i belongs to.

We can model this situation by introducing a 4-level latent variable U_i representing the latent group to which subject i belongs: $U_i \in \{11, 10, 01, 00\}$.

2.2 Graphical Models

Latent class models are based on a fundamental assumption of conditional independence and therefore they are better described using the terminology of graphical models.

Graphical models are models that describe a set of conditional independences with a graph, defined by a set of vertices, representing the variables,

and a set of edges joining the nodes such that the non-edges represent pairwise independences. (Lauritzen 1996, [17]).

Two basic types of graph used are directed and undirected graphs. In directed graphs the edges are arrows and thus the connected nodes are not on equal standing.

In this work we are dealing with *undirected graphs*, that is graphs where the edges do not have an orientation and we can regard them simply as pairs of vertices. Graphs like these are used to represent conditional dependencies: an example of this is given by the *star graphs*, where one all the nodes are connected to a single central node. This represent the case when multiple variables are independent when conditioned to a single variable.

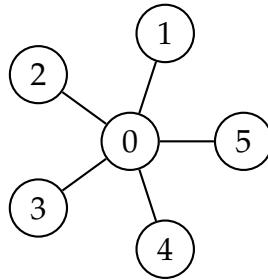


Figure 2.2.1: An example of undirected graph: the *star graph*

Another example of the use of graphs to represent dependencies is given by the *trees*, a natural extension of the star graphs to more complicate models. A notable applications of these concepts can be found in biology, specifically in phylogenetics: some structures called *phylogenetic trees* are useful to model the situation where some observed data is supposed to be generated by a complex hidden structure ([21]).

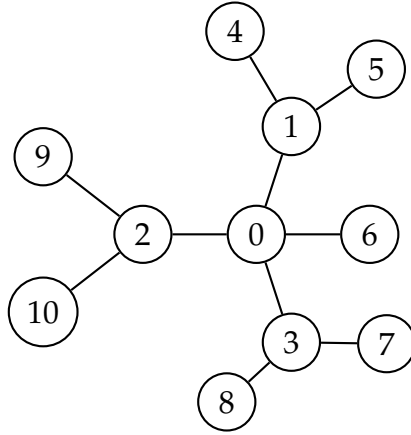


Figure 2.2.2: An example of *tree*

In order to proceed, we need the following definitions.

Definition. Let $G = (V, E)$ be a graph.

- Two vertices $u, v \in V$ are *adjacent* if $\{u, v\} \in E$, i.e. there exists an edge between them.
- The *boundary* of $v \in V$ is the set $\text{bd}(v)$ of vertices of G adjacent to v .
- G is *complete* if $E = \{\{u, v\} | u, v \in V\}$, i.e. all vertices of G are adjacent to each other.
- A *subgraph* of G is a graph $G' = (W, F)$ with $W \subseteq V$ and $F \subseteq E$. If $F = \{\{u, v\} \in E | u, v \in W\} \subseteq E$ we say that G' is the subgraph *induced* by W and we denote it by G^W .
- A subgraph of G is a *clique* if it is complete and maximal with respect to the inclusion, i.e. it is not a proper subgraph of a complete subgraph of G .
- Let $u, v \in V$. A *path* from u to v on G is a sequence $i_1 = u, i_2, \dots, i_m = v$ of vertices of G such that $\{i_j, i_{j+1}\} \in E$ for every $j \in \{1, 2, \dots, m-1\}$.

- A set of vertices of G $S \subseteq V$ *separates* two vertices $u, v \in V$ if every path on G from u to v intersect S . S separates two sets of vertices $A, B \subseteq V$ if it separates every vertex in A from every vertex in B .

We can use graphs to describe the independence structure of the joint distribution of a set X_1, \dots, X_k of random variables. To do this, we associate every variable X_i to a vertex i in a graph G .

Finally, the **complementary edge graph** (sometimes also named *missing edge graph*) of a graph $G = (V, E)$ is the graph $\overline{G} = (V, \{\{u, v\} \notin E \mid u \neq v\})$, that is, the graph with the same vertices as G having edges only between vertices that are not connected in G .

2.2.1 Markov Properties

The connection between graphs as mathematical objects and statistical models is given by the so-called undirected graph Markov properties. There are three versions of the Markov properties.

Definition. Let $\mathbf{X} = (X_1, \dots, X_k)$ be random vector of size k and $G = (V, E)$ an undirected graph with k vertices. The joint probability distribution of \mathbf{X} is said to satisfy

- the *pairwise Markov property* with respect to G if

$$X_i \perp\!\!\!\perp X_j \mid X_{V \setminus \{i, j\}} \quad \text{if } \{i, j\} \notin E;$$

- the *local Markov property* with respect to G if

$$X_i \perp\!\!\!\perp X_{V \setminus \text{bd}(i) \cup \{i\}} \mid X_{\text{bd}(i)} \quad \text{for every } i \in V;$$

– the *global Markov property* with respect to G if

$$X_A \perp\!\!\!\perp X_B | X_S \quad \text{if } S \text{ separates } A \text{ and } B \text{ in } G.$$

The following result (see Lauritzen, 1996 [17]) shows that the three Markov properties are in fact, equivalent in an important special case.

Theorem 2.2.1. *Let $\mathbf{X} = (X_1, \dots, X_k)$ be random vector of size k and $G = (V, E)$ an undirected graph with k vertices. If the joint probability distribution of \mathbf{X} is strictly positive then the three Markov properties with respect to G are equivalent.*

2.3 Log-linear Models

Let $\mathbf{X} = (X_1, \dots, X_k)$ be a discrete random vector, with variable X_i taking values in the set of non-negative integers $\{1, \dots, l_i\}$ for each $i = 1, \dots, k$: its distribution can be defined by an array (or tensor) of probabilities associated with the k -tuple of states $\{p_{i_1, \dots, i_k}\}$ and therefore it can be seen as a point in the **unit simplex**:

$$\Delta_L = \{\mathbf{p} \in \mathbb{R}^L \mid \sum p_i = 1 \text{ and } p_i \geq 0 \ \forall i = 1, \dots, L\},$$

where $L = \prod_{j=1}^k l_j$.

We shall always assume that the joint probabilities are strictly positive.

Let's assume that we have a random sample from a multinomial distribution and let \mathbf{Y} be a (vectorized) contingency table obtained by cross-classifying the frequencies from the sample according to $X_1 \dots X_k$.

In Whittaker's book on graphical models ([29]) it is shown that the natural logarithm of every discrete multivariate probability density can be written as

$$\log f_K(x) = \sum_{a \subseteq K} u_a(\mathbf{x}_a),$$

where the sum is taken over all possible subsets of $K = \{1, \dots, k\}$ and u_a are functions depending only on the values x_i for $i \in a$.

These terms are called *interactions* between variables in a . The cardinality of a is the *order* of the interaction term.

The link between log-linear expansion and conditional independence is shown in the following proposition.

Proposition 2.3.1. *If (X_a, X_b, X_c) is a partitioned multinomial random vector then $X_b \perp\!\!\!\perp X_c | X_a$ if and only if all u -terms in the log-linear expansion with one or more coordinate in b and one or more coordinate in c , are zero.*

Under the assumption that the expected values of the cells $E(\mathbf{Y})$ are non-zero, a **log-linear model** for $E(\mathbf{Y})$ is written as

$$\log(E(\mathbf{Y})) = Z\boldsymbol{\beta},$$

where $\boldsymbol{\beta}$ is a column vector of parameters and Z is a design matrix made of 0s and 1s.

The design matrix In this thesis we adopt the corner point constraints to define the design matrix. Other choices are possible but, from our viewpoint, essentially equivalent.

Let's suppose first that all variables are binary, taking values in the set $\{0, 1\}$.

The saturated model is given by $E(\mathbf{Y}) = \exp(Z\boldsymbol{\beta})$, where $Z = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}^{\otimes m}$, and \otimes is the Kronecker matrix product. This parametrization can be generalized to variables with more levels.

Example 2.3.1. With this notation, the design matrix for the saturated model for

3 binary variables is

$$Z = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

The vector of the parameters is $\boldsymbol{\beta} = (\beta_{\emptyset}, \beta_1, \beta_2, \beta_{12}, \beta_3, \beta_{13}, \beta_{23}, \beta_{123})$.

The resulting model is given by $\log(\boldsymbol{\mu}) = \log(E(Y)) = Z\boldsymbol{\beta}$. More explicitly:

$$\log \begin{pmatrix} \mu_{000} \\ \mu_{100} \\ \mu_{010} \\ \mu_{110} \\ \mu_{001} \\ \mu_{101} \\ \mu_{011} \\ \mu_{111} \end{pmatrix} = \begin{pmatrix} \beta_{\emptyset} \\ \beta_{\emptyset} + \beta_1 \\ \beta_{\emptyset} + \beta_2 \\ \beta_{\emptyset} + \beta_1 + \beta_2 + \beta_{12} \\ \beta_{\emptyset} + \beta_3 \\ \beta_{\emptyset} + \beta_1 + \beta_3 + \beta_{13} \\ \beta_{\emptyset} + \beta_2 + \beta_3 + \beta_{23} \\ \beta_{\emptyset} + \beta_1 + \beta_2 + \beta_{12} + \beta_3 + \beta_{13} + \beta_{123} \end{pmatrix}.$$

Here, $\mu_{i_1 i_2 i_3 i_4}$ is the expected value of the counts in the cell $(i_1 i_2 i_3 i_4)$.

The model from example 2.3.1 takes into account all the possible interactions between the variables. Proposition 2.3.1 provides a way to include in this framework models with fixed conditional independencies. This is done by setting some interactions to zero, and therefore by removing the corresponding

columns from the design matrix.

Example 2.3.2. If we want to impose the condition $X_1 \perp\!\!\!\perp X_2 | X_3$ in the model from example 2.3.1, we just need to set $\beta_{12} = 0$ and $\beta_{123} = 0$. The resulting design matrix Z' is obtained by removing the corresponding columns from Z :

$$Z' = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$$

We write then $\log E(\mathbf{Y}) = Z'\beta$:

$$\log \begin{pmatrix} \mu_{000} \\ \mu_{100} \\ \mu_{010} \\ \mu_{110} \\ \mu_{001} \\ \mu_{101} \\ \mu_{011} \\ \mu_{111} \end{pmatrix} = \begin{pmatrix} \beta_{\emptyset} \\ \beta_{\emptyset} + \beta_1 \\ \beta_{\emptyset} + \beta_2 \\ \beta_{\emptyset} + \beta_1 + \beta_2 \\ \beta_{\emptyset} + \beta_3 \\ \beta_{\emptyset} + \beta_1 + \beta_3 + \beta_{13} \\ \beta_{\emptyset} + \beta_2 + \beta_3 + \beta_{13} \\ \beta_{\emptyset} + \beta_1 + \beta_2 + \beta_3 + \beta_{13} \end{pmatrix}.$$

This definitions can be extended for variables with any finite number of levels. For reference, see for example Darroch et Al. (1983, [9]).

2.3.1 Hierarchical Log-linear models

A log-linear model is **hierarchical** if whenever an interaction is set to zero then all interactions between sets of variables containing the variables of that interaction are also set to zero (*hierarchical principle*).

This principle is based on the fact that if the interaction among some variables is not relevant then also the interactions containing those variables must be irrelevant.

The following proposition provides a link between the Markov properties and hierarchical log-linear models.

Proposition 2.3.2. *A discrete probability distribution is Markov with respect to an undirected graph $G = (E, V)$ if and only if in the log-linear parameterization the interaction parameters indexed by all incomplete subsets of the graph are zero.*

A hierarchical model is generated by the set of maximal non-zero interactions. If the set of generators is the set of the cliques of a graph G the model is **graphical** with respect to G .

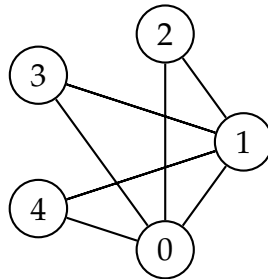


Figure 2.3.1: Graph G of the example 2.3.1

Example 2.3.3. Let $\mathbf{X} = (X_0, \dots, X_4)$ be a 5-dimensional random vector of binary variables. We consider the log-linear model on \mathbf{X} graphical with respect to

graph G in figure 2.3.1. The cliques of G are

$$\{0, 1, 2\} \quad \{0, 1, 3\} \quad \{0, 1, 4\},$$

so the model parameters are:

- β_{\emptyset} ;
- the interactions of order 1 (*principal effects*): $\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$;
- the interactions of order 2: $\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{12}, \beta_{13}, \beta_{14}$;
- the interactions of order 3: $\beta_{012}, \beta_{013}, \beta_{014}$.

We can also consider models respecting the hierarchical principle but that are not graphical by constraining some parameters to zero. For example, we can set to zero all interactions between 3 or more variables. The model then becomes

$$\begin{aligned} \log(\mu_{i_0 i_1 i_2 i_3 i_4}) = & \beta^{\emptyset} + \beta_{i_1}^{X_1} + \beta_{i_2}^{X_2} + \beta_{i_3}^{X_3} + \beta_{i_4}^{X_4} + \beta_{i_0 i_1}^{X_0 X_1} + \beta_{i_0 i_2}^{X_0 X_2} + \beta_{i_0 i_3}^{X_0 X_3} + \beta_{i_0 i_4}^{X_0 X_4} + \\ & + \beta_{i_1 i_2}^{X_1 X_2} + \beta_{i_1 i_3}^{X_1 X_3} + \beta_{i_1 i_4}^{X_1 X_4}. \end{aligned} \quad (2.1)$$

In this case, we have

$$\log(\mu_{11001}) = \beta_{\emptyset} + \beta_0 + \beta_1 + \beta_4 + \beta_{14} + \beta_{01} + \beta_{04}.$$

2.3.2 Hierarchical log-linear models with one hidden binary variable

Let $\mathbf{X} = (X_0, \dots, X_m)$ be a vector of binary variables with X_0 unobserved. We denote with Y the $2^{m+1} \times 1$ vectorized table obtained by classifying the states of

the variables and ordered in a way that the level of X_0 are changing slowest.

We assume that the entries of Y are independent Poisson variables with mean vector $\mu_Y > 0$. Again, we use a log-linear parametrization for μ_Y :

$$\log \mu_Y = Z\beta, \quad (2.2)$$

where Z is a $2^{m+1} \times q$ design matrix. We suppose now that the model is graphical according to a graph G ; q is the number of parameters in the model.

Let $\mathbf{X} = (X_0, X_1, \dots, X_k)$ be a random vector of binary variables, each taking values in the set $\{0, 1\}$.

Let ψ_0 be the parametrization of a log-linear multinomial model for the joint distribution of \mathbf{X} :

$$\begin{aligned} \psi_0 : \Omega &\rightarrow \mathbb{R}^{2^{m+1}} \\ \beta &\mapsto \exp(Z\beta). \end{aligned}$$

We denote with \mathbf{Y} the vectorized contingency table of the observed counts for a sample from a distribution from this model. We obtain the distribution μ for $E(\mathbf{Y})$ by marginalizing over the 2 possible values of X_0 :

$$\mu = \sum_{i_0=0,1} \mu_{i_0, i_1, \dots, i_k}^0, \quad (2.3)$$

where $\mu_{i_0, i_1, \dots, i_k}^0$ is the cell in the contingency table μ^0 corresponding to the states (i_0, \dots, i_k) .

We can express the marginalization in 2.3 using the matrix form:

$$\mu = L\mu_Y = L \exp(Z\beta) = \psi(\beta), \quad (2.4)$$

where $L = (1, 1) \otimes I_m = (I_m \ I_m)$ is the matrix associated with the sum over the levels of X_0 , where with I_m is the m -dimensional identity matrix with $m = 2^k$. Since we will need it later, we can now compute the jacobian matrix for this parametrization map.

Lemma 2.3.3. *The jacobian matrix for the model given by equation (2.4) is*

$$J_\psi(\boldsymbol{\beta}) = L \text{diag}(\boldsymbol{\mu}^0(\boldsymbol{\beta}))Z,$$

where $\text{diag}(\boldsymbol{\mu}^0(\boldsymbol{\beta}))$ is the matrix with $\boldsymbol{\mu}^0(\boldsymbol{\beta})$ on the main diagonal and zeros elsewhere.

Example 2.1.2 (cont.) The three employment indicators of the Job Corps Study example are plausibly associated, possibly with W12 independent of W48 given W30, and independent of the smoking habits conditional on U . Therefore, we can model this situation with a hierarchical log-linear model, like the one represented by the graph in Figure 2.3.2.

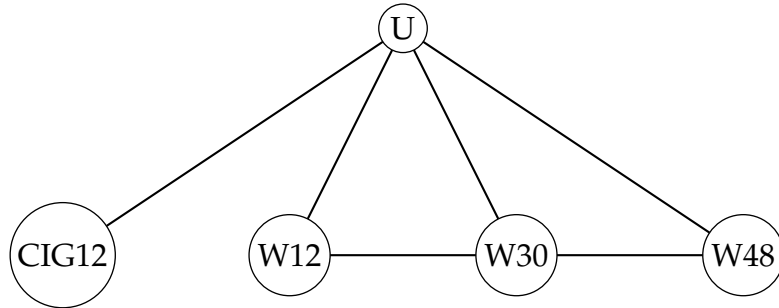


Figure 2.3.2: A possible graph for a model for the Example 2.1.2

Under suitable assumptions (see [19] for details), the latent variable U can be reduced to a binary variable. In this work we will provide some results on the identifiability of a model graphical with respect to the graph in Figure 2.3.2 and on the identifiability for alternative models as well.

Chapter 3

Identifiability

In this chapter we will briefly review the main definitions of identifiability for parametric statistical models, focusing on the cases that are more relevant to graphical and hierarchical models when latent variables are involved.

3.1 Generic and Local Identifiability

A parametric model \mathcal{M} is **identifiable** if each distribution in \mathcal{M} is given only by one parameter. This condition is also known as **strict** ([3]) or **global** ([11]) **identifiability**, in order to mark a difference with weaker identifiability conditions.

From a mathematical viewpoint, a model is strictly identifiable if the parametrization map is injective. To be true, this must hold for every point in the parameter space.

When the numbers of parameters is greater than the number of joint levels for the observed variables then identification is trivially not possible (*over-parametrization*). As next example shows, even when the number of parameters in the model is less than the number of observed levels there is no guarantee of identifiability, not even locally.

Example 3.1.1 (Goodman, 1974 [13]). Consider 4 observed binary variables X_1, X_2, X_3, X_4 , classified in a 4-way contingency table. We can model the association among

them by assuming the existence of a non-observable variable X_0 that takes values in a finite set of status $1, 2, \dots, t$ and the mutual independence of the observed variables conditional to the latent. This is an example of *Latent Class Model*.

We denote by $p_{\mathbf{i}}$ the marginal probability of the state $\mathbf{i} = (i_1, i_2, i_3, i_4)$ for the 4 observed variables, with $p_{i_j}^{i_0}$ the probability of the same state i_j for the variable X_j conditioned to the value i_0 of the latent variable X_0 and with $p_{i_1 i_2 i_3 i_4 i_0}$ the probability of the state $(i_1, i_2, i_3, i_4, i_0)$ for all the variables. With this notation the marginal probability distribution is expressed as follows:

$$p_{i_1 i_2 i_3 i_4} = \sum_{i_0=1}^t p_{i_1 i_2 i_3 i_4 i_0} = \sum_{i_0=1}^t p_{i_0} p_{i_1}^{i_0} p_{i_2}^{i_0} p_{i_3}^{i_0} p_{i_4}^{i_0}.$$

When $t = 2$, i.e. the latent variable is binary, this model is locally identifiable (see for example [3]) because the jacobian matrix in that case is in fact full rank for every choice of parameters. The parameters involved are 9 and the model is a subset of dimension 9 of the 15-dimensional unit simplex of \mathbb{R}^{16} .

As pointed out by Goodman in [13], when $t = 3$ we have instead a non-identifiable model: the number of the parameters is 14 but for a generic point in the parameter space the rank of the jacobian matrix is 13. In fact, Goodman gives 3 different estimates for the parameters that yield the same values for $p_{i_1 i_2 i_3 i_4}$.

Strict identifiability is a very strong condition, as we will see very soon, because it requires that the model is specified in the most efficient way, using only the minimal set of parameters.

This fact is particularly relevant to the present work, as we are dealing with models with latent variables: *in presence of latent variables strict identifiability is not possible*.

The reason is that the choice of the labels for the latent variable is in fact entirely arbitrary: therefore relabeling the levels of the hidden variable leads to the same probability distribution for different parameters. This is called *label swapping* issue, and shows why parametrization maps can never be injective in presence of latent variables.

Example 3.1.2 (Label swapping). Let X_1, \dots, X_k be Bernoulli random variables independent conditional to a non observable Bernoulli variable X_0 with parameter π_0 . Suppose all variables take values in $\{0, 1\}$. If we denote with $p_{x_j}^i = \mathbf{P}[X_i = 1 | X_0 = x_j]$ the conditional distribution of x_i for each level of the latent variable, the marginal distribution for the observed variables is:

$$\begin{aligned} \mathbf{P}[X_1 = x_1, \dots, X_k = x_k] &= \\ &= \pi_0 \left(\prod_{i=1}^k (p_0^i)^{x_i} (1 - p_0^i)^{(1-x_i)} \right) + (1 - \pi_0) \left(\prod_{i=1}^k (p_1^i)^{x_i} (1 - p_1^i)^{(1-x_i)} \right). \end{aligned} \quad (3.1)$$

As π_0 and $p_{x_j}^i$ range over the parameter space equation (3.1) describes all the distributions of the model. However, we note that for every parameter vector $(\pi_0, p_0^1, p_1^1, \dots, p_0^k, p_1^k)$ there is another parameter vector that gives the same distribution: $(1 - \pi_0, p_1^1, p_0^1, \dots, p_1^k, p_0^k)$. This is because the choice of the labels for the levels of the latent variable is arbitrary and inverting the parameters for the two classes gives the same distribution.

Since this issue is always present in models with latent variables, it is necessary to introduce less restricting conditions that can be applied to a wider class of models.

The need to study identifiability and their weaker notions is based on the fact that the lack of identifiability could lead to failure of standard statistical methods. The failure of injectivity can happen when the parameter space is

not smooth manifolds. It is clear that unidentified parameters cannot be consistently estimated, so identifiability is a prerequisite in statistical inference, see Example 1 in ([10]) to see the implication of the lack of identifiability in the likelihood ratio test. Thus the models with lack of global identifiability can have nonsmooth parameter spaces, so they could present difficulties for statistical inference.

When an iterative procedure is used to approximate an estimator of a parameter, different initializations can lead to multiple solutions of the estimation problem. This often corresponds to the existence of multiple parameter values giving rise to the same distributions.

For example, it is well-known that finite mixtures of Bernoulli products are not identifiable, even up to a relabelling of latent classes. However, these models are deeply used, for example in identification of bacteria ([14]).

3.1.1 Finite and local identifiability

As we have just seen, the parametrization map of a model with at least one latent variable is never 1-1: this means that for every distribution there are multiple parameters that are equivalent. If, for every distribution in the model, the set of equivalent parameters is finite we say that that model is **finitely identifiable**.

Finite identifiability also guarantees that different parameters that give the same distribution are at least isolated from each other. This property is called local identifiability, and it was introduced in 1971 by Rothenberg ([22]).

Definition. A parametric statistical model \mathcal{M} with a parametrization map $\psi : \Omega \rightarrow X$ is **locally identifiable** if ψ is locally invertible, i.e. for every $y \in \mathcal{M}$ and for every $x \in \psi^{-1}(y)$ there exists an open neighbourhood $U_y \subseteq \mathcal{M}$ of y and an open neighbourhood $U_x \subseteq X$ of x such that ψ restricted to U_x is a

diffeomorphism, i. e. an invertible differentiable function with differentiable inverse.

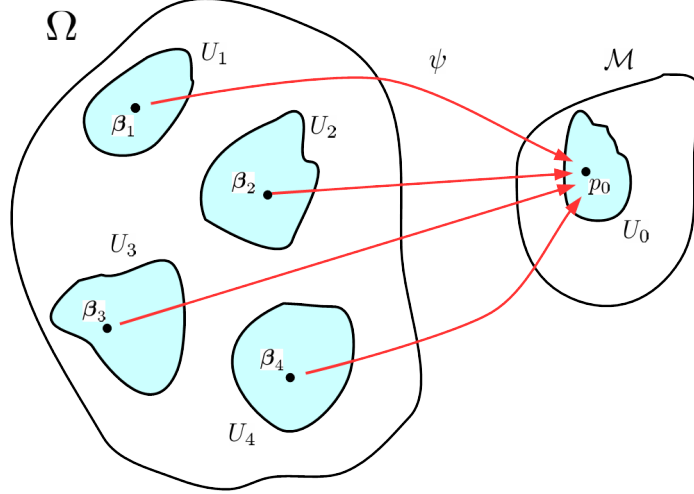


Figure 3.1.1: Local Identifiability

Local identifiability is a weaker condition than finite identifiability. However, in many cases these two conditions are equivalent.

Proposition 3.1.1. *Let \mathcal{M} be a locally identifiable parametric statistical model and let $\psi : \Omega \rightarrow \mathbb{R}^m$ be its parametrization map. If ψ is a polynomial function, then \mathcal{M} is finitely identifiable.*

Proof. Let $p \in \mathcal{M}$. We want to prove that the set $P = \psi^{-1}(p) = \{\beta \in \Omega | \psi(\beta) = p\}$ is finite.

Since \mathcal{M} is locally identifiable, P is made of isolated points and therefore is at most countable.

But P is the intersection of a finite number of zero sets of polynomials, and therefore is either finite or it contains a common component of the curves defined by these zero sets.

This implies that P is a finite set and therefore \mathcal{M} is finitely identifiable. \square

3.1.2 The Jacobian Matrix

Local identifiability implies that the sets of parameters that give the same probability distribution are made of isolated points. The main mathematical tool that we can use to assess this property is the *Inverse Function Theorem*. The proof can be found in any classical calculus textbook (e.g. Rudin's Real and Complex Analysis, [24]).

Theorem 3.1.2. *Let F be a continuously differentiable function from an open set $A \subseteq \mathbb{R}^n$ to \mathbb{R}^m , with $n \leq m$. Let $J_F(x)$ be the jacobian matrix of F for $x \in A$. Then, if $\text{rk } J_F(x) = n$ (i.e. $J_F(x)$ is full rank), there exists an open set $U_x \subseteq A$ such that $F|_{U_x}$ is 1-1.*

Now, let's suppose we have a statistical model \mathcal{M} and a parametrization map $\psi : \Omega \subseteq \mathbb{R}^q \rightarrow \mathbb{R}^m$. A sufficient condition for \mathcal{M} to be locally identifiable is that the rows of the jacobian matrix of the parametrization

$$J_\psi(\boldsymbol{\beta}) = \begin{pmatrix} \frac{\partial \psi_1}{\partial \beta_1}(\boldsymbol{\beta}) & \frac{\partial \psi_1}{\partial \beta_2}(\boldsymbol{\beta}) & \dots & \frac{\partial \psi_1}{\partial \beta_q}(\boldsymbol{\beta}) \\ \dots & \dots & \dots & \dots \\ \frac{\partial \psi_m}{\partial \beta_1}(\boldsymbol{\beta}) & \frac{\partial \psi_m}{\partial \beta_2}(\boldsymbol{\beta}) & \dots & \frac{\partial \psi_m}{\partial \beta_q}(\boldsymbol{\beta}) \end{pmatrix}$$

are linearly independent everywhere in the parameter space.

This is not, however, a necessary condition: there exist differentiable functions that are locally invertible but for which the jacobian matrix is not full rank in every point in the domain.

Example 3.1.3. The function

$$\begin{aligned} f : \mathbb{R}^2 &\rightarrow \mathbb{R}^2 \\ (x, y) &\mapsto (x^3, y) \end{aligned}$$

is infinitely differentiable. Its jacobian matrix

$$\begin{pmatrix} 3x^2 & 0 \\ 0 & 1 \end{pmatrix}$$

is rank-deficient for $x = 0$, while f is invertible (not just locally) everywhere in its domain.

For parametric models, with some regularity conditions, there is an equivalence between local identifiability of the parameters and nonsingularity of the information matrix ([22]).

3.1.3 Generic identifiability

The problem with local identifiability is that is hard to verify, because it requires that the jacobian matrix is full rank for *every* point in the parameter space. We can relax this condition by allowing some parameters to have a rank-deficient matrix, provided that the set of such points is *small*.

Definition. A parametric statistical model \mathcal{M} with a parametrization map $\psi : \Omega \rightarrow X$ is **generically identifiable** if the jacobian matrix $J_\psi(\beta)$ is full rank for almost every $\beta \in \Omega$, that is, the set of points where the rank of $J_\psi(\beta)$ is less than the number of parameters has null Lebesgue measure.

This means that the set of points at which the condition fails form a lower-dimensional algebraic subset. For example, the aforementioned finite mixtures

of multivariate Bernoulli distributions (with a fixed number of components) are generically identifiable (see Allman et al, 2009 [3]).

It is important to notice that for a polynomial model if for a single point the jacobian matrix is full rank then it is full rank everywhere in the parameter space except at most for a set of zero Lebesgue measure (see [21]).

Proposition 3.1.3. *Let \mathcal{M} be a parametric model such that the parametrization map ψ is polynomial. Then, if the rank of the jacobian matrix is full for one single parameter $\beta \in \Omega$, then it is full everywhere in Ω except at most of a subset of zero Lebesgue measure.*

This is, for example, the case of hierarchical log-linear models. The parametrization map is

$$\psi(\beta) = \exp(Z\beta). \quad (3.2)$$

By applying a parameter change $\mathbf{t}_i = \exp(\beta_i)$, the parametrization map becomes polynomial, and, since marginalization is a linear operation, also hierarchical log-linear models with latent variables can be written as image of a polynomial map.

A consequence of this fact is that to verify if a hierarchical log-linear model is generically identifiable it is sufficient to find *a single point* in the parameter space for which the rank of the jacobian matrix is full. This means that either such a model is generically identifiable or its rank is deficient everywhere.

3.2 Geometric Interpretation

Let \mathcal{M} be a parametric statistical model and $\psi : \Omega \subseteq \mathbb{R}^p \rightarrow \mathbb{R}^m$ its parametrization map.

If the jacobian matrix of this parametrization J_ψ is full rank everywhere, from the inverse function theorem \mathcal{M} is locally identifiable.

If the subset of points where J_ψ has not full rank has zero Lebesgue measure, then some distributions in \mathcal{M} can be given by a non discrete set of parameters in Ω .

Example 3.2.1. Let \mathcal{M}_G be a model that is graphical with respect to the graph G in figure 3.2.1, where X_1, X_2, X_3 and X_4 are binary observed variables and X_0 is a latent binary variable. This model is generically identifiable.

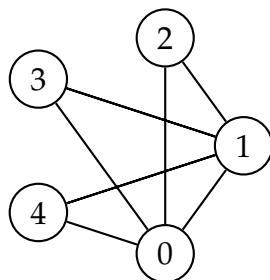


Figure 3.2.1: \mathcal{M}_G is not locally identifiable

The same distribution in \mathcal{M}_G is given by a set in the parameter space Ω that is not made by isolated points, therefore this model is not locally identifiable. We can use then the results of Stanghellini and Vantaggi (2013, [26]) to determine the set of parameters where the rank of the jacobian matrix is not full.

We will see now that, if the rank of J_ψ is not full for every point in the parameter space, \mathcal{M} is not locally identifiable.

Under the hypothesis that $\text{rk } J_\psi < p$, the null space of J_ψ is non-trivial for every $\beta \in \Omega$, that is, for every $\beta \in \Omega$ there exists a direction v_β such that $J_\psi v_\beta = \mathbf{0}$.

Let $\beta_0 \in \Omega$. We can define a curve γ in Ω passing through β_0 and tangent in every point to the direction in the null space of J_ψ :

$$\gamma(0) = \beta_0, \quad \gamma'(t) = v_{\beta(t)},$$

for $t \in [0, 1]$. Since γ is a curve in the parameter space, its image through the

parametrization map ψ is a curve on the model \mathcal{M} .

Using the chain rule, we can find the tangent vector to this new curve:

$$\begin{aligned}\Gamma'(t) &= \frac{d}{dt}\psi(\gamma(t)) = \\ &= J_\psi(\gamma(t))\gamma'(t) = J_\psi(\beta)v_\beta = \mathbf{0}.\end{aligned}$$

This implies that Γ is a constant curve in the model. We conclude that, when the model is rank-deficient everywhere, every $\beta \in \Omega$ is on a curve γ for which the parametrization is constant, and therefore the model is not locally identifiable.

3.3 Identifiability for Poisson and the multinomial models

The Log-linear parametrization can be used to define either Poisson or multinomial models for counting data. We will show now that those two cases are essentially equivalent, when we need to assess local identifiability.

Theorem 3.3.1 (Equivalence of identifiability for multinomial and Poisson models).

Let \mathcal{M} be a hierarchical model

Proof. Parametrization for the Poisson model is

$$\begin{aligned}\phi_P : \mathbb{R}^q &\rightarrow \mathbb{R}^{2^k} \\ \beta &\mapsto \exp(A\beta).\end{aligned}$$

To get the multinomial model we apply the normalization:

$$\begin{aligned}\eta : \mathbb{R}^{2^k} &\rightarrow \mathbb{R}^{2^k} \\ \mathbf{y} &\mapsto \frac{\mathbf{y}}{\sum_i y_i}.\end{aligned}$$

The multinomial model is then given by:

$$\tilde{\phi}_M = \eta \circ \phi_P.$$

We have

$$J_{\tilde{\phi}_M} \frac{\partial}{\partial \boldsymbol{\beta}} \tilde{\phi}_M(\boldsymbol{\beta}) = \frac{\partial}{\partial \mathbf{y}} \eta(\phi_P(\boldsymbol{\beta})) \times \frac{\partial}{\partial \boldsymbol{\beta}} \phi_P(\boldsymbol{\beta}).$$

Let $r_0 = \text{rk } J_{\tilde{\phi}_M}$. Since $\text{rk } \eta = 2^k - 1$ for every $\mathbf{y} \in \mathbb{R}^{2^k}$, we have $r_0 \leq \min\{q - 1, 2^k - 1\}$, so that parametrization can never be full rank.

The multinomial model is obtained as the image of the Poisson parametrization ϕ_P restricted to the set

$$X = \{\boldsymbol{\beta} \in \mathbb{R}^q \mid \sigma(\phi_P(\boldsymbol{\beta})) - 1 = 0\},$$

where

$$\begin{aligned} \sigma : \mathbb{R}^{2^k} &\rightarrow \mathbb{R} \\ \mathbf{y} &\mapsto \sum_i y_i. \end{aligned}$$

Let $\boldsymbol{\beta} \in X$ and $\boldsymbol{\omega} = \phi_P(\boldsymbol{\beta})$. We have

$$1 = \sum_i \omega_i = \sum_i \exp(A_i \cdot \boldsymbol{\beta}) = \exp(\beta_1) \sum_i \exp(A_i \cdot \boldsymbol{\beta}_0^*), \quad (3.3)$$

where $\boldsymbol{\beta}_0^* = (0, \underbrace{\beta_2, \dots, \beta_q}_{\boldsymbol{\beta}^*})$. The last equality is implied by the fact that the first column of A is made entirely of 1s.

From (3.3) we get $\beta_1 = -\log \sum_i \exp(A_i \cdot \boldsymbol{\beta}_0^*) := \gamma(\boldsymbol{\beta}^*)$. By writing $\Gamma(\boldsymbol{\beta}) = (\gamma(\boldsymbol{\beta}^*), \boldsymbol{\beta}^*)$

we can define the parametrization for the multinomial model as

$$\psi_M(\boldsymbol{\beta}^*) = \phi_P(\Gamma(\boldsymbol{\beta}^*)).$$

Using the chain rule we get the jacobian matrix for this parametrization:

$$J_{\psi_M}(\boldsymbol{\beta}^*) = J_{\phi_P}(\Gamma(\boldsymbol{\beta}^*)) \times J_{\Gamma}(\boldsymbol{\beta}^*). \quad (3.4)$$

Since

$$\text{rk } J_{\Gamma}(\boldsymbol{\beta}^*) = \text{rk} \begin{pmatrix} \frac{\partial \gamma}{\partial \boldsymbol{\beta}}(\boldsymbol{\beta}^*) \\ I_{q-1} \end{pmatrix} = q - 1$$

from (3.4) we obtain that ψ_M is full rank if and only if the intersection of the range of J_{Γ} with the null space of J_{ϕ_P} is $\{0\}$. This implies that if the Poisson parametrization is full rank then is also full rank the multinomial parametrization. □

Chapter 4

Identifiability for Hierarchical Log-linear Models with one hidden variable

A full characterization of all graphical models for which the parametrization map is full rank everywhere in the parameter space has been given by Stanghellini and Vantaggi in 2013 ([26]) in their paper on Bernoulli, for the case of one binary latent variable and any number of discrete observed variables. In this chapter we'll briefly illustrate this result and prove a generalization to the case of hierarchical models that are not graphical, when the level of the interactions is bounded to 2.

This generalization is non-trivial.

4.1 Identifiability of graphical models with one binary latent Variable

The following theorem establishes necessary and sufficient graphical conditions under which the parametrization map for a graphical model with one binary latent variable has a full rank jacobian matrix.

Theorem 4.1.1 (Stanghellini-Vantaggi, 2013). *Let \mathcal{M}_G be a graphical model over discrete variables X_0, \dots, X_m , with X_0 non-observable binary variable. Let S be the subset of the vertices of $G \setminus \{0\}$ adjacent to 0 and let $T = V \setminus (S \cup \{0\})$ the set of the*

other vertices.

Let $O = S \cup T$, the subset of the vertices of G corresponding to the observed variables. The jacobian matrix for \mathcal{M}_G is full rank everywhere in the parameter space if and only if:

- (i) the complementary of the subgraph induced by S , $\overline{G^S}$, has at least one clique of order ≥ 3 ;
- (ii) for each clique C^0 in G^S with $|C^0| > 1$ there exists a generalized identifying sequence entirely contained in S , i.e. a sequence $\{S_s\}_{s=0}^q$ of complete subgraphs of G^S such that:
 - for $s \leq q - 1$ and for all $i \in S_s$ there exists a $j \in S_{s+1}$ such that $(i, j) \notin E$;
 - $|S_{s+1}| \leq |S_s|$ for $s \leq q - 1$, $S_0 = C^0$ and $|S_q| = 1$.

Conditions (i) and (ii) of Theorem 4.1.1 together guarantee the existence of an ordering of the variables that makes the jacobian matrix block triangular. Condition (ii) is needed for all the blocks of the matrix to be invertible: if it fails it is possible to find the subspace where the rank of the jacobian matrix is not full by equating to zero the determinants of the blocks described above.

Example 4.1.1. Let G be the graph in Figure 2.3.1. The graphical model defined by G , marginalized over the latent variable X_0 , does not satisfy assumption (ii) of theorem 4.1.1, thus the jacobian matrix of the parametrization is not full rank everywhere. Cliques of G^O are:

$$C_1 = \{1, 2\}, \quad C_2 = \{1, 3\}, \quad C_3 = \{1, 4\},$$

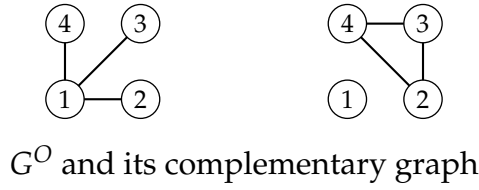
none of which admits a generalized identifying sequence.

Using the results in [26] we find that the set where identifiability breaks is the union of the subsets defined by the following linear equations:

$$\beta_{02} + \beta_{012} = 0$$

$$\beta_{03} + \beta_{013} = 0$$

$$\beta_{04} + \beta_{014} = 0.$$



Corollary 4.1.2. *If condition (i) of Theorem 4.1.1 holds then the model is generically identifiable.*

Proof. If condition (i) holds, then the space where the rank is not full can be derived from [26], where it is proved that this subspace has null Lebesgue measure.

Since this is a zero-measure set, the model is generically identifiable. \square

Example 4.1.2. The converse of this corollary is not true: there are generically identifiable models where condition (i) does fail.

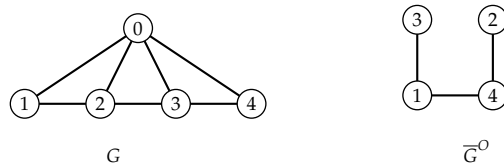


Figure 4.1.1: A model that is graphical with respect to graph G is generically identifiable even if condition (i) does not hold.

This model defined by the graphs in Figure 4.1.1 is generically identifiable, but condition (i) does not hold.

The set of parameters where the rank is not maximal can be determined:

$$\{\boldsymbol{\beta} \in \Omega \mid \beta_{01} + \beta_{012} = 0, \beta_{04} + \beta_{034} = 0\}$$

This implies that even for graphical models generic identifiability is not completely characterized.

Example 2.1.2 (cont.) Using Theorem 4.1.1 we can finally prove that the graphical model in Figure 2.3.2 we introduced to model the Job Corps Example is locally identifiable.

4.2 Non graphical Hierarchical Log-Linear Models

Theorem 4.1.1 works for models generated by cliques. Now, we will see what happens to non-graphical models, that is models obtained by constraining to 0 some parameters of a graphical model. These models are generated by a set of maximal interactions $\mathcal{A} = \{A_j\}$.

The parametrization for μ_X is now:

$$\log \mu_x = Z_{(G, \mathcal{A})} \boldsymbol{\beta}, \tag{4.1}$$

where $Z_{(G, \mathcal{A})}$ is a $2^{m+1} \times q$ design matrix defined in a way that the model is hierarchical generated by \mathcal{A} and factorizes according to the Markov property with respect to a graph G ; q is the number of parameters in the model. From now on we will write Z for $Z_{(G, \mathcal{A})}$.

Let $\psi_0 : (\mathbb{R} \setminus \{0\})^q \rightarrow \mathbb{R}^{2^{m+1}}$ be the parametrization of the joint distribution. The jacobian matrix for this function is

$$J_\psi(\boldsymbol{\beta}) = L \operatorname{diag}(\mu_x(\boldsymbol{\beta})) Z.$$

Again, we do the substitution $\mathbf{t} = \exp(\boldsymbol{\beta})$ that transforms the matrix in a matrix with polynomial entries denoted by $J_{\mathbf{t}}$ and we simplify the jacobian matrix as before.

We denote now by U_0 the family of subsets I of $\{1, \dots, m\}$ for which both \mathbf{t}_I and \mathbf{t}_{0I} are among the parameters of the model (i.e. $I \subseteq A_j$, with $A_j \in \mathcal{A}$), and we let U_1 be the set of the indices corresponding to the other parameters. Let $\mathbf{z}_1, \dots, \mathbf{z}_s$ be the columns of the matrix Z corresponding to the parameters in U_0 , and let $\mathbf{z}_{s+1}, \dots, \mathbf{z}_p$ be the columns corresponding to the parameters in U_1 . The jacobian matrix J has the following structure:

$$J = (\mathbf{z}_1, \dots, \mathbf{z}_k, D\mathbf{z}_{s+1} + \mathbf{z}_{s+1}, \dots, D\mathbf{z}_q + \mathbf{z}_q, D\mathbf{z}_1, \dots, D\mathbf{z}_s),$$

where D is the matrix with μ_x on the diagonal and zeros elsewhere.

This decomposition implies the following result.

Lemma 4.2.1. *The jacobian matrix of the parametrization (4.1) is full rank if and only if the matrix*

$$J = (\mathbf{z}_1, \dots, \mathbf{z}_k, D\mathbf{z}_{k+1} + \mathbf{z}_{k+1}, \dots, D\mathbf{z}_p + \mathbf{z}_p, D\mathbf{z}_1, \dots, D\mathbf{z}_k),$$

is full rank.

Let \mathcal{M}_G be a graphical model over variables X_0, \dots, X_k defined by a graph G , with X_0 non-observable. We denote with \mathcal{M}_G^k the hierarchical loglinear model obtained from \mathcal{M}_G by taking into account only interactions no more than k observed variables and interactions among no more than $k - 1$ observed variables and the latent one.

The following result is a generalization of Proposition 2 in [26].

Proposition 4.2.2. Let $G = (V, E)$ be an undirected graph and assume that in \overline{G}^O there exists an m -clique C , $m \geq 3$. Let $D(\beta)$ be the jacobian matrix for the hierarchical model \mathcal{M}_G^h . Let $\overline{C} = \{O \setminus C\}$ and M_1 be the sub-matrix of $D(\beta)$ formed by the rows d_i and d_{ij} , with $i \in \overline{C}$ and j such that $(i, j) \in \overline{E}$, and by the columns β_i and β_{0i} . Then, if \overline{G}^O is not connected, M_1 has rank equal to $2|\overline{C}|$ everywhere in the parameter space if and only if $h \leq 2$.

Furthermore, if G^O is connected M_1 has full rank everywhere in the parameter space.

Proof. Since \overline{G}^O is not connected, there exist 2 or more connected components. Let $\overline{G}_1 = (V_1, E_1)$ and $\overline{G}_2 = (V_2, E_2)$ be two of them. Consider any pair of complete sets $I_1 \subseteq V_1, I_2 \subseteq V_2$ in G^O . Since $(u, j) \in \overline{E}$ for any $u \in I_1, j \in I_2$, $I_1 \cup I_2$ is a complete subset of G^O .

Let $S = I_1, S'$ any complete subset of $I_1 \cup I_2$ such that $S \subseteq S'$. The 2×2 matrix formed by rows d_S and $d_{S'}$ and columns β_S and β_{0S} is

$$\begin{pmatrix} \exp(a)(1 + \exp(b)) & \exp(a + b) \\ \exp(a + a')(1 + \exp(b + b')) & \exp(a + a' + b + b') \end{pmatrix}, \quad (4.2)$$

where

- $a = \beta_\emptyset + \sum_{i \in S} \beta_i$
- $b = \beta_0 + \sum_{i \in S} \beta_{0i}$
- $a' = \sum_{i \in S', I \not\subseteq S} \delta(I) \beta_i$
- $b' = \sum_{i \in S', I \not\subseteq S} \delta(I) \beta_{0i}$,

where $\delta(I) = 1$ if I is complete and 0 otherwise. This matrix is full rank if and only if

$$\sum_{I \subseteq S', I \not\subseteq S, |I| \leq k-1} \beta_{0I} = 0. \quad (4.3)$$

Equation (4.3) defines a non-trivial subset of the parameter space if and only if $k > 2$.

Conversely, if \bar{G}^O is connected there exists an ordering of the nodes of \bar{C} such that for any $i, 1 \leq i < |\bar{C}|$, the node $j = i + 1$ is such that $(i, j) \in \bar{E}$, for $i = |\bar{C}|, j \in C$. Such ordering generates $|\bar{C}|$ distinct pairs $(i, i + 1)$. Let M_1^* be the sub-matrix of M_1 made up of the rows d_i, d_{i+1} . Then M_1^* is a $2|\bar{C}|$ -square lower-block triangular matrix with blocks M_i associated to row d_i, d_{i+1} , and columns β_i and β_{0i} . The structure of M_i is as (4.2) with $a = \beta_\emptyset + \beta_i, b = \beta_0 + \beta_{0i}, a' = \beta_j$ and $b' = \beta_{0j}$ since, by construction $(i, j) \in \bar{E}$. As $\beta_{0j} \neq 0$ by assumption, it follows that M_1 is full rank and so is M_1 . \square

4.2.1 Identification for hierarchical models with interactions up to the second order

First, we will consider models where all interactions of order higher than 2 are set to zero (so $h = 2$). In these models there is a parameter for every vertex and for every edge of the graph, so the model is defined entirely by the graph. Let G be an undirected graph with n vertices: we will denote the hierarchical model with only interactions up to the second order with the symbol \mathcal{M}_G^2 .

Example 4.2.1. With reference to graph G of example 4.1.1 we consider the model \mathcal{M}_G^2 .

The two groups of parameters are:

- $U_0 = \{\beta_\emptyset, \beta_0, \beta_1, \beta_{01}, \beta_2, \beta_{02}, \beta_3, \beta_{03}, \beta_4, \beta_{04}\};$
- $U_1 = \{\beta_{12}, \beta_{13}, \beta_{14}\}.$

This model, unlike the corresponding graphical one, is locally identifiable, since it is possible to find a maximal minor in the jacobian matrix that is invertible for every choice of the parameter.

For these models we find a complete characterization of the local identifiability.

Theorem 4.2.3. \mathcal{M}_G^2 is locally identifiable if at least one of the two following conditions holds:

- a) there exist in $\overline{G^O}$ at least one complete subgraph of order 3 or a vertex of degree at least 3;
- b) $|G^O| \geq 5$.



Figure 4.2.1: If at least one of these two graphs is a subgraph of $\overline{G^O}$, then \mathcal{M}_G^2 is locally identifiable

To prove this result we need a few propositions. The first useful property of these models is that identifiability is preserved when we add a new variable to an identifiable model.

Proposition 4.2.4. Let G be a graph and M_0 be the jacobian matrix for the log-linear parametrization of \mathcal{M}_G^2 , with $k \geq 3$. Let G_{k+1} be the graph obtained by adding a vertex v to G_0 together with edges $\{0, v\}$ and $\{j, v\}$ for $j \in W \subseteq V(G)$. Let M be the jacobian matrix for the log-linear parametrization of $\mathcal{M}_{G_{k+1}}^2$. Then, if M_0 is full rank then M is also full rank for every choice of W .

Proof. Let X_{k+1} be the variable associated with vertex v .

Since

$$\frac{\partial \mu_{(I,0)}}{\partial \beta} = 0$$

for all parameters β involving variable X_{k+1} , matrix M can be arranged in block form

$$M = \begin{pmatrix} M_0 & \mathbf{0} \\ * & \tilde{M}_{k+1} \end{pmatrix}.$$

The columns of M_0 are linearly independent so it is sufficient that \tilde{M}_{k+1} is made of linearly independent columns for M to be full rank.

We denote with d_S , with $S \subset \{1, \dots, k\}$, the row of the jacobian matrix corresponding to the derivatives of the entry μ_I where $I = (i_1, \dots, i_k)$ with $i_j = 1$ if $j \in S$ and 0 if $j \notin S$ (ex. d_{12} is the row corresponding to the derivatives of $\mu_{110\dots 0}$).

\tilde{M}_{k+1} is made by rows $d_{(I,1)}$ for every $I \in \{0,1\}^k$ and columns associated with the partial derivatives with respect to β_{k+1} and $\beta_{j,k+1}$ for every $j \in W \cup \{0\}$. Since

$$\mu_{(I,1)} = \left(\exp(\beta_{k+1} + \sum_{j \in W} \beta_{j,k+1} \delta(i_j - 1)) \right) \left(\mu_I^{(0,k)} + \exp(\beta_{0,k+1}) \mu_I^{(1,k)} \right),$$

where $\mu^{(0,k)}_I = \eta_{(0,I,0)}$ and $\mu^{(1,k)}_I = \eta_{(0,I,1)}$, we have

$$\begin{aligned} \frac{\partial \mu_{(I,1)}}{\partial \beta_{k+1}} &= \mu_{(I,1)} \\ \frac{\partial \mu_{(I,1)}}{\partial \beta_{j,k+1}} &= \mu_{(I,1)} \text{ if } j \in W, 0 \text{ otherwise} \\ \frac{\partial \mu_{(I,1)}}{\partial \beta_{0,k+1}} &= \left(\exp(\beta_{k+1} + \sum_{j \in W \cup \{0\}} \beta_{j,k+1}) \right) \mu_I^{(1,k)}. \end{aligned}$$

The columns of \tilde{M}_{k+1} are then

$$(D^{(0,k)} + D^{(1,k)})X_{\emptyset}^{(k)}, D^{(1,k)}X_0^{(k)}, (D^{(0,k)} + D^{(1,k)})X_{j_1}^{(k)}, \dots, (D^{(0,k)} + D^{(1,k)})X_{j_m}^{(k)},$$

for $j_s \in W$, where

$$\begin{aligned} & - D^{(0,k)} = \text{diag} \left(L^{(0,k)} \exp(X^{(k)} \gamma^{(k+1)}) \right) \text{ and } D^{(1,k)} = \text{diag} \left(L^{(1,k)} \exp(X^{(k)} \gamma^{(k+1)}) \right), \\ & \text{with } \gamma_v^{(k+1)} = \begin{cases} \beta_v + \beta_{v,k+1} & \text{if } v \in \{0, 1, \dots, k\} \\ \beta_{\emptyset} + \beta_{k+1} & \text{if } v = \emptyset \\ \beta_v & \text{otherwise.} \end{cases} \\ & - X_v^{(k)} \text{ is the column of } X^{(k)} \text{ associated with parameter } \beta_v. \end{aligned}$$

We notice now that \tilde{M}_{k+1} is obtained by the jacobian matrix of a latent class model with one hidden binary variable and k binary observed variables after the substitution $\gamma \rightarrow \beta$. It is known that when $k \geq 3$ this matrix is full rank if and only if the interactions γ_{0j} are not zero: since $\gamma_{0j} = \beta_{0j}$ are assumed non zero, we conclude that M is full rank. \square

From this result we deduce a sufficient condition for models with interactions up to the second order:

Proposition 4.2.5. \mathcal{M}_G^2 is locally identifiable if there exists a subgraph G_0 of G such that $\mathcal{M}_{G_0}^2$ is locally identifiable.

Proof. Let M_0 be the submatrix M made by rows d_I with $I \in \{(i_1, i_2, \dots, i_m, 0, \dots, 0) \mid (i_1, \dots, i_m) \in \{0, 1\}^m\} \subseteq \{0, 1\}^k$ and columns relative to the partial derivatives with respect to parameters involving the variables in G_0 . We define a sequence G_1, \dots, G_m of subgraphs of G such that

$$- G_m = G;$$

$$- V(G_{i+1}) = V(G_i) \cup \{v_i\}, E(G_{i+1}) = E(G_i) \cup \{\{u, v_i\} \mid \{u, v_i\} \in E(G)\}.$$

We denote with M_i the submatrix of M with rows $d_{(I,0)}$ for $I \in \{0,1\}^{|V(G_i)|}$ and columns relative to the partial derivatives with respect to parameters involving the variables in G_i for $i = 0, 1, \dots, m$. From Proposition 4.2.4 we get that if M_i is full rank then M_{i+1} is full rank, and, since M_0 is full rank, the theorem is proved. \square

Another consequence of the absence of interactions beyond the second order is that submodels of an identifiable model are also identifiable, as the following proposition shows.

Proposition 4.2.6. *If $G = (V, E)$ is a graph such that \mathcal{M}_G^2 is locally identifiable then, for every graph $\tilde{G} = (V, \tilde{E})$ obtained from G by selecting some edges $\tilde{E} \subseteq E$, $\mathcal{M}_{\tilde{G}}^2$ is locally identifiable.*

Proof. The generic element of the jacobian matrix M for the parametrization of \mathcal{M}_G^2 has the form

$$\frac{\partial \mu_I}{\partial \beta_\gamma} = \left(\exp(\mu_0 + \sum_{(u,v) \in E} \beta_{u,v} i_u i_v) \right) \left(1 + \exp(\sum_{h=1}^k \beta_{0,h} i_h) \right)$$

when $0 \notin \gamma$ and

$$\frac{\partial \mu_I}{\partial \beta_\gamma} = \left(\exp(\mu_0 + \sum_{(u,v) \in E} \beta_{u,v} i_u i_v) \right) \left(\exp(\sum_{h=1}^k \beta_{0,h} i_h) \right)$$

when $\gamma = (0, h)$ for some $h = 1, \dots, k$. The elements of a row are either constant or 0, so we can divide every row by the factor

$$\left(\exp(\mu_0 + \sum_{(u,v) \in E} \beta_{u,v} i_u i_v) \right)$$

so that the resulting matrix M^* has the same rank of the jacobian matrix of the model.

If we apply the same row operations to the jacobian matrix \tilde{M} for the model $\mathcal{M}_{\tilde{G}}^2$, we obtain a matrix \tilde{M}^* which is a submatrix of M^* . Thus, if M^* is full rank, so it is \tilde{M}^* and consequently model $\mathcal{M}_{\tilde{G}}^2$ is locally identifiable. \square

We have shown so far that identifiability is preserved both by adding a vertex or deleting some edges from a graph whose corresponding model is identifiable. To complete the classification is now sufficient to find a complete graph G such that \mathcal{M}_G^2 is identifiable:

Proposition 4.2.7. *Let G be a graph such that G^O is the complete graph with 5 vertices, like the one shown in figure 4.2.2. Then \mathcal{M}_G^2 is locally identifiable.*

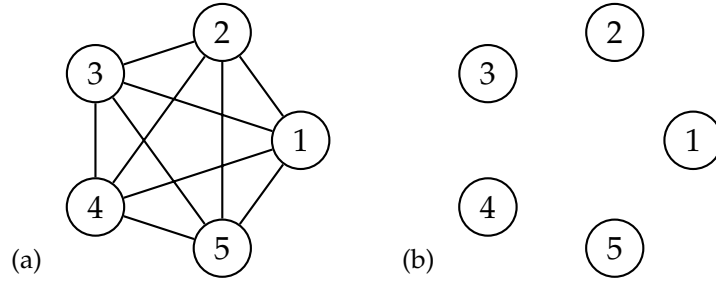


Figure 4.2.2: G^O and $\overline{G^O}$

Proof. We notice that G has no identifiable subgraph, so we can't apply Proposition 4.2.4. Let $G^{(i)}$ be the subgraph of G obtained by omitting vertex i . For every choice of the vertex i , $\mathcal{M}_{G^{(i)}}^2$ is rank-deficient in the subspace defined by equation

$$f_i = 2\beta_0 + \sum_{j \neq i} \beta_{0j} = 0. \quad (4.4)$$

Suppose \mathcal{M}_G^2 is not full rank for some vector of parameters β^* . Since equation

(4.4) must hold for every choice of i , we obtain

$$\beta_{0i}^* = -\frac{\beta_0^*}{2} \quad \forall i = 1 \dots 5. \quad (4.5)$$

But the determinant of the minor obtained by taking rows

- d_I for $I \subseteq \{1, 2, 3\}$,
- d_4 and $d_{i,4}$ for $i \in \{1, 2, 3\}$,
- d_5 and $d_{i,5}$ for $i \in \{1, 2, 3\}$,
- d_I for every I such that $|I| = 4$,
- d_{12345} ,

is zero only if $\beta_0^* = \beta_{01}^*$, implying $\beta_0^* = 0$. So β^* is outside the parameter space for our model and thus \mathcal{M}_G^2 is locally identifiable. \square

We can now prove the theorem.

Proof of Theorem 4.2.3 We classify these models by the number of observed variables:

- $|G^O| \leq 2$. None of these models is identifiable since the number of parameters is greater than the number of entries.
- $|G^O| = 3$. The latent class model with 3 observed variables is known to be identifiable (see for ex. [3]) and it is the only one with fewer parameters than number of entries.
- $|G^O| = 4$. From Proposition 4.2.5 we get that if there is a 3-clique in \overline{G}^O then \mathcal{M}_G^2 is locally identifiable, as this structure induces a full rank sub-model.

Let's suppose then that there is no 3-clique in \overline{G}^O . This means that either G has exactly 2 distinct complete components or G is connected.

- if G has exactly 2 complete components, there are 2 possible cases, shown in Figure 4.2.3.

Direct inspection of the jacobian matrix of $\mathcal{M}_{G_1}^2$ shows that in the set

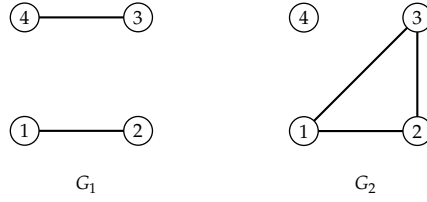


Figure 4.2.3: Two complete components

$$C_4(\Omega) = \{\beta \in \Omega \mid 2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0\},$$

the rank is 11 instead of 12.

Determinant of the submatrix of the jacobian matrix of model $\mathcal{M}_{G_2}^2$ formed by rows

- * d_I for $I \subseteq \{1, 2, 3\}$;
- * $d_4, d_{14}, d_{24}, d_{34}, d_{1234}$.

is always nonzero in the parameter space so $\mathcal{M}_{G_2}^2$ is always full rank.

- if G is connected, there are at least 3 edges. Since we are supposing that there is not a 3-clique, there must be in G at least 2 edges between 2 disjoint pairs of vertices. In this case, G has the graph G_1 from Figure 4.2.3 as subgraph. By Proposition 4.2.6, \mathcal{M}_G^2 can't be full rank everywhere.

Table 4.1 summarizes identifiability for models with 4 observed variables.

- $|G^O| \geq 5$. In this case there exists in G^O a subgraph H with 5 vertices. By propositions 4.2.7 and 4.2.6, \mathcal{M}_H^2 is locally identifiable so, by Proposition 4.2.5, \mathcal{M}_G^2 is identifiable as well. \square

It is also worth noting that all these models are at least generically identifiable and that the subset of \mathbb{R}^p where identifiability breaks is always the same linear space

$$C_4(\Omega) = \{\beta \in \Omega \mid 2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0\}, \quad (4.6)$$

because of the absence of interaction terms beyond the second order.

Actually, from Theorem 4.2.3 it follows the following characterization of generic identifiability for these models.

Corollary 4.2.8. *Let \mathcal{M}_G^2 be a hierarchical model over binary variables X_0, \dots, X_m with interaction up to the second order. Let $\mathcal{A} = \{A_i\}$ be the set of generators of \mathcal{M}_G^2 and S be the subset of the vertices j of $G^O = G \setminus \{0\}$ such that $\{j, 0\} \subseteq A_i \in \mathcal{A}$ and let $T = V \setminus (S \cup \{0\})$ the set of the other vertices.*

The jacobian matrix for \mathcal{M}_G^2 is full rank almost everywhere in the parameter space if and only if at least one of the following conditions holds:

1. $|G^S| \geq 4$;
2. $\overline{G^S}$ is a complete graph with 3 vertices.

Proof. Theorem 4.2.3 implies that all models \mathcal{M}_G^2 over X_0, X_1, \dots, X_m binary variables is locally identifiable if $m \geq 5$. When $m = 4$ from condition 2. of Theorem 4.2.3 implies that characterizes when $J_\psi^2(\beta)$ is full everywhere and when this condition fails the fullness break down in the sub-space in equation 4.6, that has zero Lebegue measure, so generic identifiability holds. When $m = 3$ just the latent class model is not overparametrized, so the thesis follows. \square

Example 2.1.2 (cont.) Using Theorem 4.2.3, we can prove that the log-linear hierarchical model defined by the graph in Figure 2.3.2 is locally identifiable when we include only interactions up to the II order.

Hierarchical models with interactions up to the second order























	G^O	$\overline{G^O}$	Identifiable	Critical set	# of par	gen. rk	crit. rk
1			YES	-	10	10	
2			YES	-	11	11	
3			Almost Everywhere	$2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0$	12	12	11
4			YES	-	12	12	
5			YES	-	13	13	
6			YES	-	13	13	
7			Almost Everywhere	$2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0$	13	13	12
8			Almost Everywhere	$2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0$	14	14	12
9			Almost Everywhere	$2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0$	14	14	13
10			Almost Everywhere	$2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0$	15	15	13
11			Almost Everywhere	$2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0$	16	16	13

Table 4.1: Identifiability for hierarchical models with 4 observed binary variables with interactions up to the second order.

The last 3 columns are, respectively: the number of parameters of the model (i.e. the dimension of the parameter space), the generic rank of the jacobian matrix for the model and rank of the jacobian matrix in the set where it is not full.

4.3 Extension to the case of observed variables with more than two levels

The aim is to extend the characterizations related to hierarchical models with interaction up to the second order over binary variables given in Theorem 4.2.3 and Corollary 4.2.8 to the case where the observed variables are not necessarily binary.

Actually, from Theorem 4.2.3 a sufficient condition follows.

Theorem 4.3.1. *Let \mathcal{M}_G^2 be a hierarchical model over variables X_0, X_1, \dots, X_m with interaction up to the second order. Suppose that the hidden variable X_0 is binary and X_i has l_i levels ($i = 1, \dots, m$). Any vertex $j \in O = \{1, \dots, n\}$ is such that $\{j, 0\} \subseteq A_i \in \mathcal{A}$. The jacobian matrix for \mathcal{M}_G^2 is full rank everywhere in the parameter space if and only if at least one of the following conditions holds:*

1. $|G^O| \geq 5$;
2. *there exist in $\overline{G^O}$ at least one m -clique C with $m \geq 3$ or a vertex of degree 3 or a vertex of order 3.*

Proof. Assume firstly that all the variables are binary except the variable X_1 , which has three levels. Partition β into three subsets:

- $\beta^a = \{\mu, \beta_0\}$;
- $\beta^b = \{\beta_1^1, \beta_{0,1}^1, \beta_i, \beta_{0,i}, \beta_{1,i}^1, \beta_{i,j} : i, j = 2, \dots, m, i \neq j\}$, corresponding to the non-zero interaction terms for value in $\{0, 1\}$ of the observable random variables;
- $\beta^c = \{\beta_1^2, \beta_{0,1}^2, \beta_{1,i}^2 : i = 2, \dots, m\}$ containing all other parameters.

After ordering in a way such that the X_1 variable is running the slowest, the $J_\psi^2(\beta)$ matrix has the following structure:

$$J_\psi^2(\beta) = \begin{bmatrix} D(\beta^a) & D(\beta^b) & 0_{2^n \times |\beta^c|} \\ D^*(\beta^a) & 0_{2^{(n-1)} \times |\beta^b|} & D^*(\beta^c) \end{bmatrix}$$

where $[D(\beta^a) \mid D(\beta^b)]$ is the sub-matrix of the derivatives of β^a and β^b .

It has full rank if the observable variables are at least 5 or the condition 2. of Theorem 4.2.3 holds.

Note that by construction, $D^*(\beta^c)$ has a similar structure of the sub-matrix of $D(\beta^b)$ formed by the last $2^{(n-1)}$ rows and all columns. Therefore $D^*(\beta^c)$ is full rank if condition 2. of Theorem 4.2.3 holds or again the observable variables are at least 5.

Proof of the theorem for X_1 having l_1 levels follows straightforwardly.

By a similar argument, extension to a generic number of levels of the X_i variables, $i \in O$, follows. \square

Actually, from Theorem 4.3.1 the following condition assuring generic identifiability follows.

Corollary 4.3.2. *Let \mathcal{M}_G^2 be a hierarchical model over variables X_0, \dots, X_m with interaction up to the second order. Suppose that the hidden variable X_0 is binary. Let $\mathcal{A} = \{A_i\}$ be the set of generators of \mathcal{M}_G^2 and S be the subset of the vertices j of $G^O = G \setminus \{0\}$ such that $\{j, 0\} \subseteq A_i \in \mathcal{A}$ and let $T = V \setminus (S \cup \{0\})$ the set of the other vertices.*

The jacobian matrix for \mathcal{M}_G^2 is full rank almost everywhere in the parameter space if at least one of the following conditions holds:

1. $|G^S| \geq 4$;

2. $\overline{G^S}$ is a complete graph with 3 vertices.

Chapter 5

Identifiability for models with higher order of interactions

So far we have characterized all the models with binary variables that are full rank everywhere in the parameter space when we have limited the order of the interactions to 2. We have found that, in this class, every model with at least 5 observed binary variables and one binary latent variable is locally identifiable. A characterization of generically identifiable models is also provided. The case of higher interactions is a bit more complicated, as we are also considering triple interactions between 2 observed variables and the latent one.

5.1 Generic Identifiability for models with III order interactions

Now we are going to consider models with interactions of order 3 or more, some of the previous results cannot be extended to this case since the conditions are no longer true. Concerning proposition 4.2.6, we observe that the graph G_1 is



Figure 5.1.1: Proposition 4.2.6 cannot be extended to higher orders.

obtained from G_2 by deleting edge $\{3, 4\}$: however, $\mathcal{M}_{G_2}^3$ is locally identifiable,

while $\mathcal{M}_{G_1}^3$ is not (not even generically). This induces to look for further investigation. In fact, Proposition 4.2.6 is fundamental for proving the main result concerning local identifiability Theorem 4.3.1 and the main result concerning generic identifiability, Corollary 4.3.2.

In the following we analyze the different situations that can arise by looking at the number of observed variables.

5.1.1 Models with 4 observed variables

First, we review the conditions we have given on models with interactions of order 2 when we include higher orders.

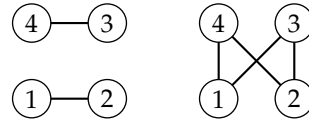


Figure 5.1.2: Direct and complementary graphs for the only model with 4 observed variables without a 3-clique in the complementary graph that is not overparametrized when we include III order interactions.

This model is the only graphical model with 4 observed variables without a 3-clique in the complementary graph that is not overparametrized. Its generic rank is 12, while its parameter space is 14-dimensional.

It can be shown that the last two columns of the jacobian matrix J , corresponding to the partial derivatives with respect to β_{34} and β_{034} , are a linear combination of the first 12 columns: the coefficients can be calculated solving the linear system $J^0 \lambda = J^1$.

After the variable change $\mathbf{t} = \exp(\boldsymbol{\beta})$, the solution of this system is:

$$\begin{pmatrix} \alpha & -\beta \\ -\alpha & \beta \\ 0 & -(\mathfrak{t}_{01} - 1)\beta \\ \frac{(\mathfrak{t}_{01}-1)\alpha}{\mathfrak{t}_{01}} & 0 \\ 0 & -(\mathfrak{t}_{02} - 1)\beta \\ \frac{(\mathfrak{t}_{02}-1)\alpha}{\mathfrak{t}_{02}} & 0 \\ 0 & (\mathfrak{t}_{01} + \mathfrak{t}_{02} - \mathfrak{t}_{01} \mathfrak{t}_{02} \mathfrak{t}_{012} - 1)\beta \\ (\mathfrak{t}_{01} \mathfrak{t}_{012} + \mathfrak{t}_{02} \mathfrak{t}_{012} - \mathfrak{t}_{01} \mathfrak{t}_{02} \mathfrak{t}_{012} - 1)\alpha & 0 \\ (\mathfrak{t}_{03} - 1)\alpha & 0 \\ 0 & -\frac{(\mathfrak{t}_{03}-1)\beta}{\mathfrak{t}_{03}} \\ (\mathfrak{t}_{04} - 1)\alpha & 0 \\ 0 & \frac{\mathfrak{t}_{04}-1}{\mathfrak{t}_{04}-\mathfrak{t}_{04}^2 \mathfrak{t}_{034} + \mathfrak{t}_{03} \mathfrak{t}_{04}^2 \mathfrak{t}_{034} - \mathfrak{t}_{03} \mathfrak{t}_{04} \mathfrak{t}_{034}} \end{pmatrix},$$

where

$$\alpha = \frac{1}{\mathfrak{t}_{03} + \mathfrak{t}_{04} - \mathfrak{t}_{03} \mathfrak{t}_{04} \mathfrak{t}_{034} - 1}$$

$$\beta = \frac{1}{\mathfrak{t}_{03} \mathfrak{t}_{034} + \mathfrak{t}_{04} \mathfrak{t}_{034} - \mathfrak{t}_{03} \mathfrak{t}_{04} \mathfrak{t}_{034} - 1}$$

The submatrix made by rows d_I , for $I = \emptyset, 1, 2, 12, 3, 13, 23, 123, 4, 14, 34, 1234$, has the maximal rank (12).

We recall that, when we focus on hierarchical models with interactions up to the second order, the above graph (as well as all the graphs with 4 observed variables including that graph) correspond to a model that is generic identifiable but the rank of the jacobian matrix is not full everywhere. The rank of that model is not full in a subspace of null Lebesgue measure and in that space is 11.

This shows that the situation, as aforementioned, is much more intriguing.

The peculiarity of this graph is that the complementary graph contains a cycle with 4 variables. In the following we will emphasize that this structure has a central role in the identifiability or better in the lack of generic identifiability..

5.1.2 Models with 5 observed variables

As aforementioned in the previous section, models with complementary graph over the observed variables with cycles of length 4 have a central role when interactions of order 3 are considered.

Model A

We consider a hierarchical model with interaction up to the III order over 5 observed variables, that is represented in Figure 5.1.3). The peculiarity of this complementary graph is that we can take in consideration any group of 4 observed variables and the subgraph (of the complementary one) is a cycle. Since it will be important in the following of this work, we name it Model A.

In the following result we study the generic rank of the jacobian matrix of this hierarchical model with interactions up to the III order that has 21 parameters.

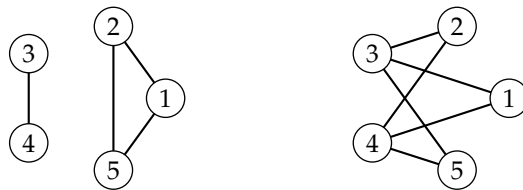


Figure 5.1.3: Direct and complementary graph of the observed variables for Model A

Proposition 5.1.1. *Generic rank of Model A is 20.*

Proof. All submodels obtained by deleting a vertex are rank deficient.

Removing vertices 3 or 4 gives a model equivalent to the overparametrized graphical model in Figure 5.1.4 that has a jacobian matrix with 17 linearly de-

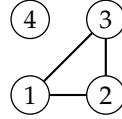


Figure 5.1.4: Overparametrized graphical submodel of model A

pendent columns.

Removing vertices 1, 2 or 3 gives a model equivalent to the graphical model in Figure 5.1.5, which is also rank deficient.

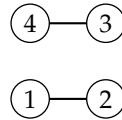


Figure 5.1.5: Non-identifiable graphical submodel of model A

□

The submatrix made by rows and columns has rank 20. Since the number of parameters of the model is 21 and the generic rank of the jacobian matrix is 20, the model is not generic identifiable.

Model B

Now the aim is to study models having more interactions (among the observed variables) than the model A. Obviously, adding edges in the directed graph, the corresponding edges in the complementary graph need to be deleted, so the resulting complementary graphs have a group of 4 observed variables such that the subgraph (of the complementary one) has no 4-cycle.

Proposition 5.1.2. *Model B is generically identifiable.*

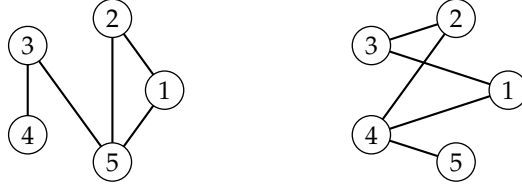


Figure 5.1.6: Direct and complementary graph of the observed variables for model B

Proof. Let J be the jacobian matrix for this model:

$$J = \begin{pmatrix} M_H & \mathbf{0}_{16 \times 7} \\ J_{21} & J_{22} \end{pmatrix}$$

Here the ordering of the variables is 23451, so M_H is the jacobian matrix for the model defined by the subgraph H of G generated by vertices 2, 3, 4 and 5.

Since \mathcal{M}_H^3 is generically identifiable, M_H is full rank except in a set C_H of zero measure. Since this submodel is graphical, we can use the result from [26] to find such set: $C_H = \{\beta | \beta_{01} + \beta_{012} = 0, \beta_{04} + \beta_{034} = 0\}$.

We want now to construct a submatrix of J full rank almost everywhere. First we consider the submatrix of J obtained by taking the first 15 rows from M_H (all except for d_{2345}), together with rows d_1, d_{14}, d_{124} and d_{125} . Let M_{19} be the matrix formed by these rows and columns correspondent to all the parameters in M_H except for together with $\beta_1, \beta_{01}, \beta_{12}$ and β_{012} .

M_{19} is a lower block triangular square matrix of order 19 and it's full rank when $\beta \notin C_H \cup \{\beta | \beta_{05} - \beta_{04} + \beta_{015} + \beta_{025} = 0\}$. \square

Model B is full rank almost everywhere. However, at least one of its submodels is rank deficient everywhere. This implies that proposition 4.2.6 for models with at most interactions of order 2 does not hold when interactions of the III order (or higher) are considered.

5.2 Theorem

Starting from the above considerations we can state the main result concerning hierarchical models with interactions up to the III order.

Theorem 5.2.1. *When $|O| \geq 4$, \mathcal{M}_G^3 is generically identifiable if and only if at least one of the two following conditions holds:*

- a) there exists at least a submodel defined by a subgraph of G with 4 vertices that is neither overparametrized nor there is a 4-cycle in \overline{G}^O .*
- b) $|O| \geq 6$.*

Proof. Let's suppose that there exists a submodel \mathcal{M}_h^3 that satisfies condition a), for some subgraph H .

Without loss of generality we can suppose that the variables in that submodel are labeled 1, 2, 3 and 4.

From Theorem 4.1.1, this submodel is full rank. If $|O| = 4$, $G^O = H$, and the result is proven. If $|O| = 5$, the jacobian matrix of \mathcal{M}_G^3 can be written as a block triangular matrix.

$$\begin{pmatrix} J_H & \mathbf{0} \\ * & M_2 \end{pmatrix},$$

Where J_H is the jacobian matrix of \mathcal{M}_H^3 . Therefore, we only need to prove that M_2 is full rank almost everywhere. By applying the same transformation in we used in Proposition 4.2.6, we find that this submatrix is a submatrix of the jacobian matrix of a locally identifiable model, with new parameters γ . This matrix is full rank except in the set $\gamma = 0$, which, in the original parameter space, has a Lebesgue measure of zero. Finally, if $|O| \geq 6$, the model is a submodel of the complete model with 6 observed variables, which is graphical and generically identifiable, and therefore it is at least generically identifiable. \square

5.2.1 Consequences of the characterization of generic identifiability for models with interactions up to III order

In [26] the rank of jacobian matrix of graphical models have been studied. A characterization(Theorem 4.1.1) of the models with full rank everywhere is provided. In Section 4.1 the main implications of Theorem 4.1.1 have been addressed in order to study generic identifiability. In particular, when condition (i) of Theorem 4.1.1 holds the graphical model is generic identifiable. Now the aim is to fully classify the cases where condition (i) for graphical models fails:

- 2 complete components in G^O with at least 2 elements each \rightarrow **NO GENERIC IDENTIFIABILITY.**

This is due to the fact that in the complementary graph over n observed variables the following situations can occur:

- when $n = 4$ there is a 4-cycle in the complementary graph;
- when $n > 4$ the two cliques in the directed graph can have $i \geq 2$ and $n - i \geq 2$ variables, respectively, with at least one inequality strict. In this case, any subgraph over 4 observed variables is overparametrized or corresponds to a complementary graph with 4 observed variables having a 4-cycle.

This implies that the model is not generically identifiable.

- 1 connected component in G^O where every subgraph with 4 vertices is either a 4-cycle in \overline{G}^O or induces an overparametrized submodel \rightarrow **NO GENERIC IDENTIFIABILITY**

The considerations for this situation are similar to the previous one, so generic identifiability cannot hold.

- 1 connected component in G^O such that there are 4 observed variables with the submodel not overparametrized and with complementary graph that is not a 4-cycle. → **GENERIC IDENTIFIABILITY**

The generic identifiability follows when the interactions are up to the III order by previous results.

5.3 Generic Identifiability for models with interactions of order 4 or more

We have seen that some of the fundamental results in the previous chapter fail in the presence of interactions of order 3 between the latent variable and two observed variables. However, in this section we will use some of those propositions to prove a slightly weaker result.

The first step is to prove a generalization of Proposition 4.2.5

Proposition 5.3.1. *Let \mathcal{M}_G^k be a hierarchical log-linear model, with $k \geq 3$, on a random Poisson vector $\mathbf{X} = (X_0, \dots, X_m)$, with X_0 binary latent variable. Let v be a vertex in $G = (V, E)$ and $G^{(v)} \subset G$ be the subgraph of G obtained by removing v .*

We define, given v , a new graph $G_2 = (V_2, E_2)$ in this way:

- $V_2 = V \setminus \{v\}$;
- $E_2 = \{(i, j) \in V_2 \times V_2 \mid \text{the subgraph induced by } \{i, j, v\} \text{ in } G \text{ is complete}\}.$

If there exists a vertex v such that $\mathcal{M}_{G^{(v)}}^k$ is generically full rank and $\mathcal{M}_{G_2}^{k-1}$ is always full rank, then \mathcal{M}_G^k is generically full rank.

Proof. Let M be the jacobian matrix of \mathcal{M}_G^k . We can reorder the vertices in V in a way that v is the last one.

M has a natural block structure:

$$\begin{pmatrix} M_1 & \mathbf{0} \\ * & M_2 \end{pmatrix}.$$

Submatrix M_1 is exactly the jacobian matrix of model $\mathcal{M}_{G(v)}^k$, therefore is full rank almost everywhere in the parameter space.

The matrix M_2 concerns with interactions involving v , since \mathcal{M}_G^k the interactions have order up to k , the interactions obtained by deleting v are of order up to $k - 1$ and since the rank of $\mathcal{M}_{G_2}^{k-1}$ is full everywhere, it follows that M_2 is full rank. \square

Now we have a rule for proving if a hierarchical log-linear model is generically full rank:

- we look in G for submodels with one vertex less that are generically full rank;
- if such a model is found, we check if the model defined in the proposition (that has a maximal level of interaction smaller than the original model) is one full rank everywhere.

The aim is to provide a further sufficient condition assuring local identifiability.

Theorem 5.3.2. *Let \mathcal{M}^h be a hierarchical model over the binary variables X_0, X_1, \dots, X_m with interaction up to the order h . Suppose that \mathcal{M}^h is Markovian with respect to the graph G .*

Let \mathcal{M}_1^h be hierarchical model, sub-model of \mathcal{M}^h over $X_0, X_{i_1}, \dots, X_{i_d}$ with $i_j \in \{1, \dots, m\}$ ($j = 1, \dots, d$) such that \mathcal{M}^h restricted to the variables $X_0, X_{i_1}, \dots, X_{i_d}$ coincides with \mathcal{M}_1^h .

If the jacobian matrix for the submodel is full rank everywhere in the parameter space and for any maximal interaction in \mathcal{M}^h there exists a generalized identifying sequence

$\{I_i\}_{i=0,\dots,q}$, then the jacobian matrix for \mathcal{M}^h is full rank everywhere in the parameter space.

Proof. Let J_G be the jacobian matrix for \mathcal{M}^h and J_{G_1} the jacobian matrix for \mathcal{M}_1^h . Since J_{G_1} is full rank everywhere in the parameter space, there exists a square matrix D_C of J_{G_1} of dimension equal to the columns of J_{G_1} having full rank everywhere in the parameter space.

In order to prove that J_G is full rank everywhere in the parameter space, note that

$$J_G = \begin{pmatrix} J_{G_1} & 0 \\ A & B \end{pmatrix}$$

so it is necessary to show that also B has a square submatrix of dimension the number of columns of B with full rank everywhere.

If $h = 2$ the result follows since $m \geq 4$. If $h \geq 3$ \bar{G}^0 is connected (see Proposition 5.2.1). Since for any clique in G there is a generalized identifying sequence and by putting $\bar{C} = \{O \setminus C\}$ where $C = \{1, \dots, m_1\}$ there exists an ordering (see the algorithm in Appendix A of [26]) of the vertices of \bar{C} such that for any i , $1 \leq i < |\bar{C}|$, the vertex $j = i + 1$ is such that $(i, j) \in \bar{E}$; for $i = |\bar{C}|$, $j \in C$. Such ordering generates $|\bar{C}|$ distinct pairs $(i, i + 1)$.

This implies that B can be so written

$$\begin{pmatrix} J_{G_1} & 0 & 0 \\ B_1 & D_{\bar{C}} & 0 \\ B_2 & B_3 & P \end{pmatrix}.$$

$D_{\bar{C}}$ is formed by the rows d_i and $d_{\{i,j\}}$, with $i \in \bar{C}$ and j such that $(i, j) \in \bar{E}$, and by the columns β_i and $\beta_{\{0,i\}}$; while P is the matrix related to interaction I_k of order k with $k \geq 2$ among the observed variables and of order $(k + 1)$ among k observed variables and the hidden variable X_0 .

Then, $D_{\bar{C}}$ has rank equal to $2|\bar{C}|$ everywhere in the parameter space if \bar{G}^O is connected (see Proposition 5.2.1).

Concerning P^G , note that matrix P is formed by the matrices $M_{k,r}$ of order $2(q+1)$ formed by the rows d_{I_s} and $d_{\{V, I_s\}}$, $V \subseteq I_{s+1}$, and by the columns associated to β_{I_s} and $\beta_{\{0, I_s\}}$, $s \in \{0, \dots, q\}$, where $I_{k,r}$ is a complete subgraph of G^0 with k the number of the observed variables and $\{I_s\}_{s=0, \dots, q}$ is the generalized identifying sequence (that exists since $J_\psi^G(\beta)$ is full rank everywhere by Theorem 4.1.1).

Now, concerning $J_\psi^h(\beta)$ the matrix P is formed by the matrices $M_{k,r}$ with $k \leq h-1$ and the matrices $M'_{h,r}$ formed by the row $d_{I_{h,r}}$ and the column $\beta_{I_{h,r}}$ where again $I_{h,r}$ is a complete graph in G^0 . $\beta_{I_{h,r}}$ and the rows $d_{I_{h,r}}$ coincides with the a sub-matrix of P^G . Concerning P , a row, and therefore a column, cannot be chosen twice in $M_{k,r}$ or $M'_{h,r}$ matrix, since a generalizing sequence with $I_s \neq I_{s'}$ can be obtained (see Remark 3 in [26]). By ordering the rows and columns according to the sequence of $\{I_s\}_0^{q+1}$, the matrix $M_{k,r}$ and $M'_{h,r}$ is seen to be lower block triangular. The blocks of $M_{k,r}$ are N_0, \dots, N_q where N_s is formed by the rows d_{I_s} and $d_{\{V, I_s\}}$ with $V \subset I_{s+1}$ by the columns associated to β_{I_s} and $\beta_{\{0, I_s\}}$. Therefore N_s is as in

$$\begin{pmatrix} e^a(1 + e^b) & e^{a+b} \\ e^{a+a'}(1 + e^{b+b'}) & e^{a+a'+b+b'} \end{pmatrix} \quad (5.1)$$

Then, $\text{rank}(M_{k,r}) = \sum_{s=0}^q \text{rank}(N_s)$ and is full if and only if the blocks are full rank, that is if the rank of each block is equal to 2. Since $J_\psi^G(\beta)$ has full rank, all sub-blocks of N_s have rank equal to 2.

While the block of $M'_{h,r}$ is $N_{h,r}$ formed by the row $d_{I_{h,r}}$ and the column associated to $\beta_{I_{h,r}}$ and has full rank since its entry is $e^a(1 + e^b)$ and cannot be zero.

Therefore for each s there exists a full rank block N_s and the square submatrices $M_{k,r}$, $k \leq h-1$, $M'_{h,r}$ are full rank everywhere in the parameter space. \square

The condition assuring local identifiability in Theorem 5.3.2 is based on the condition (ii) of Theorem 4.1.1, that characterizes full rank jacobian matrix for graphical models. This implies that in the case of not graphical models condition (i) of Theorem 4.1.1 can be replaced by another condition as in Theorem 5.3.2. Actually, this condition in Theorem 5.3.2 consists in the full rank of a restriction of the hierarchical model over some observed variables. Thus, this condition is a weak form of condition (i) in Theorem 4.1.1.

Theorem 5.3.2 implies the following result concerning with generic identifiability.

Theorem 5.3.3. *Let \mathcal{M}^h be a hierarchical model over the binary variables X_0, X_1, \dots, X_m with interaction up to the order h . Suppose that \mathcal{M}^h is Markovian with respect to the graph G .*

Let \mathcal{M}_1^h be hierarchical model, sub-model of \mathcal{M}^h over $X_0, X_{i_1}, \dots, X_{i_d}$ with $i_j \in \{1, \dots, m\}$ ($j = 1, \dots, d$) such that \mathcal{M}^h restricted to the variables $X_0, X_{i_1}, \dots, X_{i_d}$ coincides with \mathcal{M}_1^h .

If the jacobian matrix for the submodel is full rank everywhere in the parameter space, then \mathcal{M}^h is generically identifiable.

Proof. From the proof of Theorem 5.3.2 it follows that if there is for some interaction in \mathcal{M}^h no generalized identifying sequence the jacobian matrix is not full in a space of null measure since just some block $M_{k,r}$ degenerates. Then, the generic rank is full. \square

Some relevant results concerning with locally and generic identifiability have been provided, however for hierarchical models with interactions up to order

$k \geq 4$ the characterizations of generic or local identifiable models is not completed, there are some situations that need to be deepened further.

This problem has been studied even by means of simulations, even if the size of jacobian matrices increases exponentially with the number of observed variables, and so computational problems arise.

These investigations lead to the following conjecture concerning the maximal order of interactions and the number of observed variables.

Conjecture 1. \mathcal{M}_G^k is generically identifiable if $k \leq \frac{|G|}{2}$.

So far, we have verified this conjecture using a MATLAB script (available from the author upon request) that computes the generic rank of the jacobian matrix for log-linear hierarchical models.

A comprehensive collection of all possible topologies for graphs with at most 10 vertices has been listed using the software *nauty*.

5.4 Behaviour of the model near the critical points

We have seen that when G^O is the graph in figure 5.4.1 then in the set

$$\{\boldsymbol{\beta} \in (\mathbb{R} \setminus \{0\})^p \mid 2\beta_0 + \beta_{01} + \beta_{02} + \beta_{03} + \beta_{04} = 0\}$$

the parametrization is rank deficient by 1. Let ψ be the parametrization map for

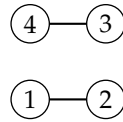


Figure 5.4.1: Graph of the observed variables for the example.

\mathcal{M}_G^2 and $\boldsymbol{\beta}_0 \in C_4(\Omega)$. The jacobian matrix $J_\psi(\boldsymbol{\beta}_0)$ is a 16×12 matrix with rank

11, so there is a vector $\mathbf{v} \in \mathbb{R}^{12}$ that generates its null space. Then, the curve defined by

$$\begin{aligned}\gamma : \mathbb{R} &\rightarrow \mathcal{M}_G^2 \\ t &\mapsto \psi(\beta_0 + t\mathbf{v})\end{aligned}$$

has one critical point for $t = 0$, when the curve passes through β_0 . In figure 5.4.2 is shown a plot of the 16 components of γ in a neighbourhood of $t = 0$.

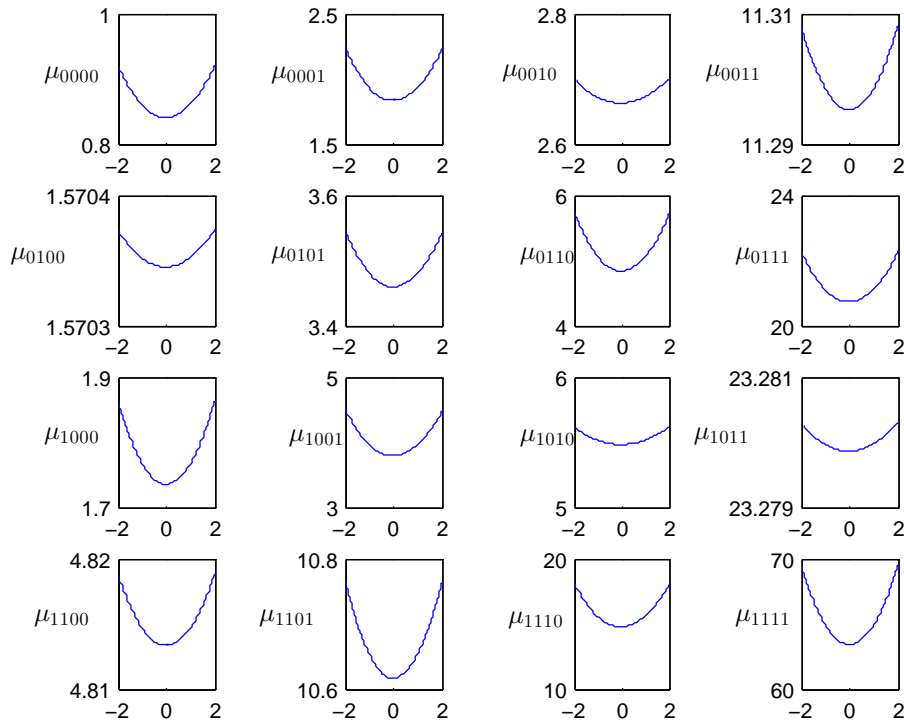


Figure 5.4.2: Behaviour of the model in a neighbourhood of a critical point

Bibliography

- [1] P.S. Albert, L.E. Dodd (2004). A cautionary note on the robustness of latent class models for estimating diagnostic error without a gold standard. *Biometrics*, 60(2), 427-35.
- [2] A. Agresti, (2013) *Categorical data analysis*, 3rd edn. Wiley, New York
- [3] E. Allman, C. Matias, and J. Rhodes (2009). Identifiability parameters in latent structure models with many observed variables. *The Annals of Statistics*, 37, 3099- 3132.
- [4] E. Allman, J. Rhodes, E. Stanghellini and M. Valtorta (2014). Parameter Identifiability of Discrete Bayesian Networks with Hidden Variables. *J. Causal Infer.* 1-17.
- [5] D. Bartholomew, M. Knott, I. Moustaki (2011). *Latent Variable Models and Factor Analysis: A Unified Approach*. Wiley Series in Probability and Statistics.
- [6] D. Bartholomew (2004). *Measuring Intelligence: Facts and Fallacies*. Cambridge University Press.
- [7] J. Burghardt, S. McConnell, P. Schochet, T. Johnson, M. Gritz, S. Glazerman, J. Homrighausen (2001). *Job Corps Work? Summary of the National Job Corps Study*. Document No. PR01-50, Mathematica Policy Research, Inc., Princeton, NJ

- [8] E. A. Catchpole, B. J. T. Morgan (1997). Detecting parameter redundancy. *Biometrika*, 84, 187-196
- [9] J. Darroch, T. Speed (1983). Additive and multiplicative models and interaction. *Annals of Statistics*, 11, 724-738.
- [10] M. Drton (2009). Likelihood ratio tests and singularities. *The Annals of Statistics*. 37(2), 979-1012.
- [11] R. Foygel, M. Drton, S. Sullivant (2009). Global identifiability of linear structural equation models. *Annals of Statistics* 39(2), 865-886.
- [12] Z. Gilula (1983). Latent conditional independence in two-way contingency tables: A diagnostic approach. *British Journal of Mathematical and Statistical Psychology*, 1983, 36(1), pp. 114-22.
- [13] L. A. Goodman (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- [14] Gyllenberg, M., Koski, T., Reilink, E. and Verlaan, M. (1994). Nonuniqueness in probabilistic numerical identification of bacteria. *J. Appl. Probab.* 31 542-548.
- [15] T. Kocka, N.L. Zhang (2002). Dimension correction for hierarchical latent class models. in *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI2002)*.
- [16] T. C. Koopman, O. Reiersol (1950). The Identification of Structural Characteristics 18th Conference on Uncertainty in Artificial Intelligence (UAI-02) *Ann. Math. Statist.* Volume 21, Number 2, 165-181.
- [17] S.L. Lauritzen (1996). *Graphical Models*. Oxford: Oxford University Press.

- [18] B.D. McKay, A. Piperno (2013). Practical graph isomorphism, II. *Journal of Symbolic Computation*, 60:94112.
- [19] F. Mealli, B. Pacini and E. Stanghellini. (2014). Identification of principal causal effects using secondary outcomes. In Carpita M., Brentari E., Qanari E. *Advances in Latent Variables. Studies in Theoretical and Applied Statistics*. Springer-Verlag.
- [20] J. Neyman (1937). Remarks on a Paper by E. C. Rhodes," *Journal of the Royal Statistical Society*, 100.
- [21] L. Pachter, B. Sturmfels (2005). *Algebraic Statistics for Computational Biology*. Cambridge University Press.
- [22] T. J., Rothenberg (1971). Identification in parametric models. *Econometrica* 39 577591.
- [23] Rubin, D.B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics*, 6, 34-58
- [24] W. Rudin (1997), *Real And Complex Analysis*, 3rd ed. McGraw-Hill, Inc. New York, NY, USA.
- [25] E. Stanghellini (1997). Identification of a single-factor model using graphical Gaussian rules. *Biometrika*, 84, 2414.
- [26] E. Stanghellini, B. Vantaggi (2013). On the identification of discrete concentration graph models with one hidden binary variable. *Bernoulli*, 19(5A), 1920-1937.
- [27] P. Vicard (2000). On the identification of a single-factor model with correlated residuals. *Biometrika*, 87, 199205.

- [28] S. Watanabe (2009). Algebraic Geometry and Statistical Learning Theory, Monographs on Applied and Computational Mathematics 25, Cambridge University Press.
- [29] J. Whittaker (1990). Graphical Models in Applied Multivariate Statistics. Wiley.